



INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(51) International Patent Classification 7 : C12N 15/90	A1	(11) International Publication Number: WO 00/63410 (43) International Publication Date: 26 October 2000 (26.10.00)
(21) International Application Number: PCT/US00/09782 (22) International Filing Date: 12 April 2000 (12.04.00) (30) Priority Data: 09/293,303 16 April 1999 (16.04.99) US (63) Related by Continuation (CON) or Continuation-in-Part (CIP) to Earlier Application US 09/293,303 (CON) Filed on 16 April 1999 (16.04.99) (71) Applicants (for all designated States except US): MASSACHUSETTS INSTITUTE OF TECHNOLOGY [US/US]; 77 Massachusetts Avenue, Cambridge, MA 02139 (US). ALBERT EINSTEIN COLLEGE OF MEDICINE OF YESHIVA UNIVERSITY [US/US]; 1300 Morris Park Avenue, Bronx, NY 10461 (US). (72) Inventors; and (75) Inventors/Applicants (for US only): BOUHASSIRA, Eric [FR/US]; 123 Farragut Avenue, Hastings on Hudson, NY 10706 (US). LEBOULCH, Philippe [FR/US]; Unit 729, 197 Eighth Street, Charlestown, MA 02129 (US).	(74) Agents: REMILLARD, Jane, E. et al.; Lahive & Cockfield, LLP, 28 State Street, Boston, MA 02109 (US). (81) Designated States: AE, AG, AL, AM, AT, AU, AZ, BA, BB, BG, BR, BY, CA, CH, CN, CR, CU, CZ, DE, DK, DM, DZ, EE, ES, FI, GB, GD, GE, GH, GM, HR, HU, ID, IL, IN, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MA, MD, MG, MK, MN, MW, MX, NO, NZ, PL, PT, RO, RU, SD, SE, SG, SI, SK, SL, TJ, TM, TR, TT, TZ, UA, UG, US, UZ, VN, YU, ZA, ZW, ARIPO patent (GH, GM, KE, LS, MW, SD, SL, SZ, TZ, UG, ZW), Eurasian patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European patent (AT, BE, CH, CY, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE), OAPI patent (BF, BJ, CF, CG, CI, CM, GA, GN, GW, ML, MR, NE, SN, TD, TG). Published With international search report. Before the expiration of the time limit for amending the claims and to be republished in the event of the receipt of amendments.	

(54) Title: METHODS AND COMPOSITIONS FOR TRANSFORMING CELLS USING SITE-SPECIFIC RECOMBINATION

(57) Abstract

Methods and compositions for transforming cells, resulting in efficient and stable site-specific integration of transgenes, are disclosed. Transformation is achieved by providing an acceptor DNA containing two inverted lox sequences and a donor DNA containing the same two inverted lox sequences, and then contacting the acceptor and donor DNA with a recombinase (e.g., Cre or Flp) which causes recombination at the lox sequences contained in the DNAs. Prior to recombination, the acceptor DNA is preferably integrated into the genome of a cell, such as an embryonic stem cell or a fertilized egg. The acceptor DNA optionally may further contain a negatively selectable marker to allow for screening of cells which have undergone the desired site-specific recombination (e.g., DNA cassette exchange).

FOR THE PURPOSES OF INFORMATION ONLY

Codes used to identify States party to the PCT on the front pages of pamphlets publishing international applications under the PCT.

AL	Albania	ES	Spain	LS	Lesotho	SI	Slovenia
AM	Armenia	FI	Finland	LT	Lithuania	SK	Slovakia
AT	Austria	FR	France	LU	Luxembourg	SN	Senegal
AU	Australia	GA	Gabon	LV	Latvia	SZ	Swaziland
AZ	Azerbaijan	GB	United Kingdom	MC	Monaco	TD	Chad
BA	Bosnia and Herzegovina	GE	Georgia	MD	Republic of Moldova	TG	Togo
BB	Barbados	GH	Ghana	MG	Madagascar	TJ	Tajikistan
BE	Belgium	GN	Guinea	MK	The former Yugoslav Republic of Macedonia	TM	Turkmenistan
BF	Burkina Faso	GR	Greece	ML	Mali	TR	Turkey
BG	Bulgaria	HU	Hungary	MN	Mongolia	TT	Trinidad and Tobago
BJ	Benin	IE	Ireland	MR	Mauritania	UA	Ukraine
BR	Brazil	IL	Israel	MW	Malawi	UG	Uganda
BY	Belarus	IS	Iceland	MX	Mexico	US	United States of America
CA	Canada	IT	Italy	NE	Niger	UZ	Uzbekistan
CF	Central African Republic	JP	Japan	NL	Netherlands	VN	Viet Nam
CG	Congo	KE	Kenya	NO	Norway	YU	Yugoslavia
CH	Switzerland	KG	Kyrgyzstan	NZ	New Zealand	ZW	Zimbabwe
CI	Côte d'Ivoire	KP	Democratic People's Republic of Korea	PL	Poland		
CM	Cameroon	KR	Republic of Korea	PT	Portugal		
CN	China	KZ	Kazakhstan	RO	Romania		
CU	Cuba	LC	Saint Lucia	RU	Russian Federation		
CZ	Czech Republic	LI	Liechtenstein	SD	Sudan		
DE	Germany	LK	Sri Lanka	SE	Sweden		
DK	Denmark	LR	Liberia	SG	Singapore		
EE	Estonia						

METHODS AND COMPOSITIONS FOR TRANSFORMING CELLS USING SITE-SPECIFIC RECOMBINATION

Government Funding

5 This invention was made with government support under grants HL38655, HL 554350, DK54071-020 and HL07556, all of which were awarded by the NIH. The United States government has certain rights in the invention.

Background of the Invention

10 The success of gene therapy techniques depends largely on the ability to achieve a combination of stable chromosomal integration and high-level, regulated expression of transferred genes. Regulated gene expression is most easily achieved by means of large DNA fragments containing extensive cis-acting regulatory regions. For example, gene therapy for β -globin disorders may require high-level, position-independent expression
15 of extended gene and LCR sequences.

 Many current techniques allow efficient transient transfection of cells *in vitro* and *in vivo* with large DNA fragments. However, subsequent chromosomal integration is very inefficient. To overcome low levels of integration, retroviral vectors which integrate very efficiently in permissive cells can be used. However, such vectors are
20 greatly limited by constraints of size and sequence composition.

 There are also many other techniques available for stable integration of transgenes in mammalian cells (Kriegler, M. (1990) *Gene Transfer and Expression. A Laboratory Manual*, Stockton Press, New York); and (Wolf, J.A. (1994) *Gene Therapeutics: Methods and Application of Direct Gene Transfer*, Birkhauser, Boston)..
25 However, these methods result in integration at random chromosomal locations of an uncontrolled number of transgene copies that express at levels that generally cannot be predicted or reproduced with precision because of position-effects. The inability to control the site of integration, the number of integrated copies and the level of expression of transgenes has impeded progress in studies of both gene expression and
30 the physiological effects of transgenes.

Systems which can perform site-specific chromosomal integration efficiently therefore have wide utility. The first site-specific chromosomal integrations in mammalian cells were based on integration of a single Lox or FRT site on a chromosome followed by trapping of rare integration events (O'Gorman et al. (1991) *Science* 251:1351-1355; and Sauer, B. (1994) *Current Opinion in Biotechnology* 5:521-527). These pioneering methods had three limitations: 1) they were quite inefficient, 2) the entire plasmid was integrated, and 3) a positive selectable marker was left in the chromosome after the integration. The low efficiency of these methods is due to the reversibility of the recombination reaction: after integration the transgene is re-excised if the two identical Lox or FRT sites that flank the transgene recombine with each other. Since the excision reaction is intra-molecular while the insertion reaction is inter-molecular, excisions are favored.

US 4,959,317 discloses the use of Cre-Lox site-specific recombination to achieve gene transfer in eukaryotic cells (Sauer et al. (1993) *Methods in Enzymology* 225: at 898). The target site of the CRE recombinase is a 34 bp sequence that consists of two inverted 13 bp CRE-binding sites separated by an eight base spacer within which the recombination occurs (Hoess, et al. (1984) *Proc. Nat. Acad. Sci. (USA)* 81:1026-1029).

Additional site-specific DNA recombination systems which provide more efficient and stable integration of transgene sequences into genomic DNA, preferably without the use of a positively selectable marker, would be greatly beneficial.

Summary of the Invention

The present invention provides methods and compositions for achieving efficient and stable site-specific DNA recombination using a recombinase/lox system, such as the Cre/lox system or the FLP/frt system. In one embodiment, the method comprises contacting a recombinase (e.g., Cre or FLP) with (a) an acceptor vector comprising two incompatible lox sequences, L1 and L2, and (b) a donor vector comprising a selected DNA flanked by the L1 and L2 sequences, or sequences which are compatible with the L1 and L2 sequences, thereby causing transfer of the selected DNA from the donor vector into the acceptor vector by recombination at the compatible lox sequences. In a

preferred embodiment, the acceptor vector is a retroviral vector or an adeno-associated vector.

In another embodiment, the invention provides a method of transforming a cell with a selected DNA comprising, in any order, the steps of introducing into the cell an
5 acceptor vector which integrates into the genome of the cell, the acceptor vector comprising two incompatible lox sequences, L1 and L2, (b) introducing into the cell a donor vector comprising the selected DNA flanked by the L1 and L2 sequences, or sequences which are compatible with the L1 and L2 sequences, and (c) contacting L1 and L2 with a recombinase, such as Cre or Flp, thereby causing transfer of the selected
10 DNA from the donor vector into the acceptor vector. The recombinase can be introduced into the cell in the form of a protein or a gene encoding the protein.

In another embodiment, the invention provides a vector selected from the group consisting of retroviral vectors and adeno-associated vectors comprising two incompatible lox sequences, L1 and L2.

15 In another embodiment, the invention provides a method of achieving site-specific recombination by providing a donor DNA comprising two inverted lox sequences, and an acceptor DNA comprising the same two inverted lox sequences contained in the donor DNA, and then contacting the donor and acceptor DNA with a recombinase (e.g., Cre or Flp). Preferably, the acceptor DNA is integrated into the
20 genome of a host cell prior to contact with the recombinase (e.g., by homologous recombination), so that recombination results in site-specific genomic integration of a desired transgene or other polynucleotide. In another preferred embodiment, the donor DNA is present in excess of the acceptor DNA. Suitable lox sequences comprise the nucleotide sequences of SEQ ID NO:1, SEQ ID NO:2 and inversions thereof. The
25 donor and acceptor DNAs can further contain selectable markers which are generally positioned between the two inverted lox sequences to help select for cells which have undergone the desired recombination.

The methods and compositions of the invention can be used in methods of *in vivo* and *In vitro* gene transfer (e.g., gene therapy) to cause efficient and stable site-
30 specific (targeted) integration of transgene sequences. By controlling the site or position of integration of a transgene into the genome of a cell, expression levels can be

predictably determined. For example, in cells, the invention can be used to produce desired proteins (e.g., drugs) by insertion of transgenes at pre-selected chromosomal locations where expression of the transgene will be high. Similarly, the invention can be used to develop "designer proteins" by insertion of multiple versions of a gene or DNA (e.g., related variants) at the same locus to test the various versions of proteins produced in a context in which the proteins are all produced in the same amount. The invention can also be used to study and to identify genetic elements that control position effects.

In addition, the invention can be used *in vivo* to create transgenic mammals and/or plants. For example, animal models of human disease can be generated, particularly if multiple genes have to be expressed at well regulated levels. Animals and/or plants can be created which contain target lox sequences (e.g., inverted or incompatible) at chromosomal locations that are not subject to position effects or to desired position effects for directing expression of a gene of interest. This allows for the generation of animal models and/or plants with, for example, higher resistance to disease or improved physical/functional characteristics.

Brief Description of the Figures

Figure 1 is a schematic illustration of Cre/lox mediated gene transfer. Lox A and Lox B are mutually incompatible lox sites which are unable to recombine with each other in the presence of Cre recombinase.

Figure 2 is an illustration of acceptor and donor vectors containing selectable marker genes and incompatible lox sequences. Following pre-integration of the acceptor vector into a genome, the donor vector undergoes site-specific Cre-mediated recombination with the acceptor vector at the lox sequences, resulting in stable DNA exchange between the vectors

Figure 3 is an illustration of a Cre expression vector and a control Cre expression vector.

Figure 4 is a schematic illustration of recombinase-mediated DNA cassette exchange using either incompatible exchange sequences (e.g., lox L1 and L2) or inverted exchange sequences (e.g., lox L1 and 1L). **Panel A** shows how site-specific chromosomal integration can be performed by exchanging a pre-integrated chromosomal cassette (cas1) with a plasmidic cassette (cas 2), using incompatible lox sequences. Recombination through the L1 site leads to plasmid integration; recombination through L2 leads to resolution of the array thus created. The end result of the recombination reaction is therefore an exchange of cassette 1 by cassette 2. This exchange is stable because the cassette is flanked by Lox sites that cannot recombine with each other.

10 Symmetrical integration through L2 and resolution through L1 also lead to the cassette exchange. After integration the resolution step is in competition with an excision step that is the reverse of the integration step. **Panel B** shows how selection for the loss of the HYTK cassette can be used to identify clones which have undergone site-specific chromosomal recombination. **Panel C** shows site-specific chromosomal recombination using inverted Lox sites: The L2 Lox site is replaced by an inverted L1 site (1L). Excision products are eliminated because recombination between Lox sites located in cis leads to inversion of the HYTK gene rather than its excision (upper recombination). Gancyclovir selection therefore lead predominantly to the identification of exchange products (lower recombination). Exchange products in both possible orientations are

20 obtained. Only one of several exchange pathways is depicted. **Panel D** shows FACS-based site-specific chromosomal recombination. A plasmid containing the CMV-EGFP reporter gene inserted outside of the exchange cassette is co-transfected (with a CRE expression plasmid) and cells that express the highest amount of GFP are sorted. Among those cells, exchange frequency is relatively high. No selectable marker on the

25 chromosome is required before or after the exchange.

Figure 5 shows Southern blots confirming site-specific chromosomal recombination in MEL and ES cells. **Panel A** is an autoradiogram illustrating the exchange of the HYTK gene with the HS234 β GFP cassette at two loci in MEL cells.

30 Lanes 1 to 6 in the left panel, and lanes 1 to 7 in right panel, contain genomic DNA from Gan^R resistant clones digested with EcoRV and Bgl II and probed with a fragment

encompassing the EGFP coding sequence. Two bands (6.1 and 5.1 at RL4 or 7.9 and 7.2 kb at RL5) were observed per locus demonstrating insertion in both possible orientations. B= BglII; X= EcoRV or BglII.

Panel B is an autoradiogram illustrating the exchange of the HYTK gene with the HS234 β GFP cassette at two loci in ES cells. Lanes 1 to 7 in the left panel contain genomic DNA from Gan^R resistant clones digested with HindIII and probed with a fragment encompassing the EGFP coding sequence. The right panel is a blot of the left panel stripped and re-probed with a fragment encompassing the CMV promoter. As expected, clones in both orientations were obtained. **Panel C** shows recombinase-mediated DNA cassette exchange using no selectable marker. Left panel, lanes 1 to 3, and right panel, lanes 1 and 2, contain genomic DNA digested with BclI and probed with a fragment encompassing the human β -globin promoter; lane C contain control DNA. As expected, all exchanges were in the same orientation. Bc= BclI. **Panel D** shows random (uncontrolled) integration of Cassette CMV-EGFP: Lanes 1 to 9 contain genomic DNA digested with AflII probed with the EGFP coding sequence. Bands of all sizes and intensity characteristic of uncontrolled integration are shown.

Figure 6 is a chart showing expression levels of GFP at pre-selected locus and at a random locus: GFP expression of nine independent clones with the CMV-EGFP cassette integrated at 9 uncontrolled random loci (left panel) or at the same pre-selected locus (right panel) was analyzed by FACS. The y-axis represents the mean linearized GFP fluorescence. The results show that site-specific integration considerably diminishes the variability of expression typically observed in stable transfection studies.

25 Detailed Description of the Invention

The present invention provides methods and compositions for causing efficient site-specific DNA recombination, for example, in cells or fertilized eggs. The advantages over current cell transformation techniques provided by the invention include highly efficient and stable integration of large DNA sequences, such as transgene sequences, into chromosomal DNA, without the need for a selectable marker in the integrated DNA. In one embodiment, the method of the invention comprises

contacting a recombinase, such as Cre, with (a) an acceptor DNA (e.g., vector) comprising two incompatible lox sequences, L1 and L2, and (b) a donor DNA (e.g., vector) comprising a selected DNA flanked by the L1 and L2 sequences, or lox sequences which are compatible with the L1 and L2 sequences, thereby causing transfer
5 of the selected DNA from the donor vector into the acceptor vector by recombination at the compatible lox sequences.

In another embodiment, the method of the invention comprises contacting a recombinase, such as Cre or Flp, with (a) a donor DNA comprising two inverted lox sequences, and (b) an acceptor DNA comprising the same two inverted lox sequences
10 contained in the donor DNA. Preferably, the acceptor DNA is integrated into the genome of a host cell (e.g., by homologous recombination), prior to contact with the recombinase so that recombination results in site-specific genomic integration of a desired transgene or other polynucleotide.

The term "site-specific recombination," refers to DNA transfer from a donor
15 DNA or vector to an acceptor DNA or vector.

The term "lox sequence" refers to a nucleotide sequence which undergoes recombination (e.g., DNA cross-over and exchange) when catalyzed by a recombinase, such as Cre, Flp or another member of the Int family of recombinases (Argos et al. (1986) *EMBO J.* 5: 433). Suitable lox sequences include, for example, the lox
20 sequences recognized by Cre recombinase, and the frt sequences recognized by Flp recombinase.

The term "recombinase" refers to any recombinase capable of catalyzing a site-specific recombination at a lox site. Suitable recombinases include, for example, Cre recombinase (Sauer et al. (1993) *Methods in Enzymology* 225: 898) and Flp recombinase
25 (Buchholz et al. (1996) *Nucl. Acids Res.* 24:4256-4262; Buchholz et al. (1998) *Nat. Biotechnol.* 16:657-662).

The term "incompatible lox sequences" refers to two or more lox sequences (referred to herein as L1 and L2) which differ from one another and, therefore, can not undergo recombination with one another. For example, lox sequences can be rendered
30 incompatible if their nucleotide sequences differ by even one nucleotide, particularly in

their spacer regions. In contrast, the term "compatible lox sequences" refers to two or more lox sequences which can recombine when catalyzed to do so by a recombinase.

The term "inverted lox sequences" refers to two or more substantially identical lox sequences (referred to herein as L1 and 1L) which are positioned in the reverse orientation with respect to one another so that recombination between the L1 and 1L sequences within the same (e.g., acceptor) DNA (e.g., intrachromosomal recombination) leads to an inversion of the intervening sequence, as opposed to an excision (see Figure 4). "Inverted lox sequences" of the invention recombine with the same lox sequences (e.g., contained within a separate donor DNA) at the same frequency (i.e., specific activity) because they are made up of identical or substantially identical nucleotide sequences. Overall, these features provide a site-specific recombination system which results in (1) substantially equal amounts of intermolecular recombination between identical L1 or 1L sequences (e.g., contained on separate acceptor and donor DNAs), and (2) substantially no excision from intramolecular recombination between inverted L1 and 1L sequences (e.g., contained within the same acceptor or donor DNA). As shown schematically in Figure 4, this allows for highly (up to 100%) efficient and stable recombinase-mediated DNA integration.

The terms "acceptor DNA" and "acceptor vector," are used interchangeably herein and refer to any DNA or vector which, preferably, is capable of integrating into the genome of a cell. For example, the acceptor DNA or vector can be of viral origin, such as a retroviral vector or adeno-associated vector. Generally, the acceptor DNA or vector contains an exchange cassette (i.e., DNA which is replaced by DNA from the donor vector) and can also, optionally, contain a selectable (e.g., negative) marker gene.

The terms "donor DNA" and "donor vector" are used interchangeably herein and refer to any DNA or vector (e.g., circular plasmid DNA) containing DNA which is transferred to the acceptor DNA or vector via a recombinase-mediated exchange reaction. Generally, the donor DNA or vector comprises plasmid DNA and, optionally, also can contain a selectable (e.g., positive) marker gene.

The method of the present invention utilizes a recombinase-mediated exchange reaction which takes place between identical or compatible (i.e., able to recombine with one another) lox sequences. The efficient exchange of DNA between identical or

compatible lox sequences enables transfer of DNA from the donor to the acceptor vector, which each contain identical or compatible lox sites (see Figure 4). However, once transferred from donor to acceptor vector (i.e., intermolecular transfer), the transferred DNA is stabilized or "locked" into place. In one embodiment, this is
5 achieved by using incompatible lox sequences (e.g., L1 and L2) within the acceptor vector which prevent intramolecular exchange and excision of the transferred DNA. In another embodiment, this is achieved by using identical inverted lox sequences (e.g., L1 and 1L) within the acceptor vector which recombine with each other in reverse so that the transferred DNA is inverted within the vector without excision of the transferred
10 DNA. Therefore, the transferred DNA is integrated in a highly efficient and stable manner.

In addition to effective and stable DNA exchange reactions, the methods of the present invention take advantage of highly efficient integration vectors, such as retroviral vectors, adeno-associated vectors, or vectors encoding retroviral integrases, for
15 use in integrating acceptor DNA into the genome of a cell. The studies described herein demonstrate that such vectors are compatible for use with site-specific DNA transfer systems, such as recombinase/lox systems.

The invention also employs an excess of donor DNA relative to acceptor DNA so that intermolecular exchange leading to insertion of donor DNA into (e.g.,
20 genomically integrated) acceptor DNA occurs at a higher frequency than intramolecular exchange (e.g., within acceptor DNA) leading to excision of donor DNA.

Overall, the site-specific recombination system of the invention provides a means for highly efficient and stable DNA transfer which can be used, for example, in methods of gene therapy, gene expression control, and transgenic applications. For
25 example, as described herein, the methods and compositions of the present invention can be used to achieve a 100 to 10,000 fold increase in transgene integration and expression compared to random integration which occurs in the absence of site-specific recombination. The invention further provides a way to achieve site-specific recombination without the incorporation of selectable markers into the genome
30 following exchange.

Accordingly, in another embodiment, the invention provides a method of transforming a selected cell or tissue, such as a mammalian cell (e.g., embryonic stem cell) or fertilized egg, with a desired DNA. The method comprises the steps of (a) introducing into the cell or tissue an acceptor vector comprising two incompatible lox sequences, L1 and L2, (b) introducing into the cell a donor vector comprising the selected DNA flanked by the L1 and L2 sequences, or lox sequences which are compatible with the L1 and L2 sequences, and (c) contacting L1 and L2 with a recombinase, such as Cre, to cause transfer of the selected DNA from the donor vector into the acceptor vector (by way of an exchange reaction between the compatible lox sequences). While not essential, the acceptor vector is preferably introduced into the cell prior to introduction of the donor vector, so that the acceptor vector has integrated into the host genome prior to DNA exchange with the donor vector.

Alternatively, selected cells and/or tissues can be transformed by (a) introducing into the cell or tissue an acceptor DNA comprising two inverted lox sequences; (b) introducing into the cell or tissue a donor DNA comprising the same two inverted lox sequences contained in the acceptor DNA; and (c) contacting the donor and acceptor DNA with a recombinase. The recombinase can be introduced into the cell or tissue in the form of an exogenous protein or in the form of a gene encoding the recombinase which is expressed in the cell or tissue following transfection. Suitable lox sequences again include those having the nucleotide sequences shown in SEQ ID NOS: 1 and 2, as well as inversions thereof.

The acceptor DNA can be any DNA (e.g., plasmid vector) capable of being taken up by cells and integrating into genomic DNA. Suitable acceptor DNAs include viral vectors which transfect cells directly, such as recombinant retroviruses, adenovirus, adeno-associated virus, and herpes simplex virus-1. A prerequisite for the use of retroviruses is to ensure the safety of their use, particularly with regard to the possibility of the spread of wild-type virus in the cell population. The development of specialized cell lines (termed "packaging cells") which produce only replication-defective retroviruses has increased the utility of retroviruses for gene therapy, and defective retroviruses are well characterized for use in gene transfer for gene therapy purposes (for a review see Miller, A.D. (1990) *Blood* 76:271). Thus, recombinant retrovirus can be

constructed in which part of the retroviral coding sequence (*gag*, *pol*, *env*) has been replaced by nucleic acid encoding a mutated subunits of the mALDH of the invention rendering the retrovirus replication defective. The replication defective retrovirus is then packaged into virions which can be used to infect a target cell through the use of a helper virus by standard techniques. Protocols for producing recombinant retroviruses and for infecting cells *in vitro* or *in vivo* with such viruses can be found in Current Protocols in Molecular Biology, Ausubel, F.M. et al. (eds.) Greene Publishing Associates, (1989), Sections 9.10-9.14 and other standard laboratory manuals. Examples of suitable retroviruses include pLJ, pZIP, pWE and pEM which are well known to those skilled in the art. Examples of suitable packaging virus lines for preparing both ecotropic and amphotropic retroviral systems include ψ Crip, ψ Cre, ψ 2 and ψ Am. Retroviruses have been used to introduce a variety of genes into many different cell types, including neural cells, epithelial cells, endothelial cells, lymphocytes, myoblasts, hepatocytes, bone marrow cells, *in vitro* and/or *in vivo* (see for example Eglitis, et al. (1985) *Science* 230:1395-1398; Danos and Mulligan (1988) *Proc. Natl. Acad. Sci. USA* 85:6460-6464).

Another suitable acceptor vector is an adenovirus-derived vector. The genome of an adenovirus can be manipulated such that it encodes and expresses a gene product of interest but is inactivated in terms of its ability to replicate in a normal lytic viral life cycle. See for example Berkner et al. (1988) *BioTechniques* 6:616; Rosenfeld et al. (1991) *Science* 252:431-434; and Rosenfeld et al. (1992) *Cell* 68:143-155. Suitable adenoviral vectors derived from the adenovirus strain Ad type 5 dl324 or other strains of adenovirus (e.g., Ad2, Ad3, Ad7 etc.) are well known to those skilled in the art.

Yet another viral vector system useful as the acceptor vector is the adeno-associated virus (AAV). Adeno-associated virus is a naturally occurring defective virus that requires another virus, such as an adenovirus or a herpes virus, as a helper virus for efficient replication and a productive life cycle. (For a review see Muzyczka et al. *Curr. Topics in Micro. and Immunol.* (1992) 158:97-129). It is also one of the few viruses that may integrate its DNA into non-dividing cells, and exhibits a high frequency of stable integration (see for example Flotte et al. (1992) *Am. J. Respir. Cell. Mol. Biol.* 7:349-356; Samulski et al. (1989) *J. Virol.* 63:3822-3828; and McLaughlin et al. (1989) *J.*

Virol. 62:1963-1973). Vectors containing as few as 300 base pairs of AAV can be packaged and can integrate.

Other viral vector systems that may be used as the acceptor vector in the methods of the present invention include herpes virus, vaccinia virus, and several RNA viruses.

5 Alternatively, the acceptor DNA can be targeted for delivery to selected cells (see e.g., U.S. 5,166,320) via cell-specific targeting ligands. Once taken up by cells, the acceptor DNA can be integrated into genomic DNA either randomly or selectively, for example, by homologous recombination as is well known in the art (see e.g., 5,614,396).

The acceptor DNA also can be introduced, for example, by microinjection and/or
10 homologous recombination, into a host cell, embryonic stem cell, tissue, or fertilized egg. The donor DNA can be any DNA (e.g., circular DNA) capable of being taken up by cells, either *in vivo* or *in vitro*, and capable of carrying the desired transfer (i.e., donor) DNA sequence which, for example, encodes a therapeutic gene. Suitable donor vectors include cosmids or DNA plasmids, such as recombinant bacterial or eukaryotic
15 plasmids. The donor vector can be introduced into the host cell either *in vivo* or *in vitro* using a variety of known methods. For *in vitro* delivery, suitable methods include direct injection of the plasmid (e.g., microinjection), CaPO_4 precipitation, electroporation, cationic lipofection, or use of artificial viral envelopes. For *in vivo* delivery, suitable methods include intravenous, intraperitoneal and intramuscular injection of the vector.

20 The vector can also be targeted for delivery to selected cells (see e.g., US 5,166,320).

In a preferred embodiment of the invention, the donor DNA is introduced into the host cell in excess of the acceptor DNA to increase the frequency of DNA
integration relative to excision. Integration is an intermolecular event, whereas excision is an intramolecular event. Therefore, excision naturally occurs more often than
25 integration. To tilt the balance toward integration, the amount of donor DNA can be increased to increase the number of intermolecular exchange reactions which occur in the host cell.

In another embodiment, the acceptor and/or donor DNA also includes a selectable marker which enables the identification of clones which have undergone site-
30 specific recombination resulting in an exchange of acceptor DNA for donor DNA. For example, the acceptor DNA can include a negatively selectable marker gene to identify

cells which have lost the ability to produce a certain protein (e.g., which have lost resistance to a drug or antibiotic). Similarly, the donor DNA can include a positively selectable marker to identify cells which have incorporated the donor DNA into their genome via recombinase-mediated exchange. The marker gene can be promoterless so
5 that it will only be expressed when integrated into the acceptor vector containing a promoter to drive its expression.

The donor and acceptor vectors each contain at least two incompatible lox sequences ("L1 and L2") so that intramolecular recombination can not occur. At the same time, the locks sequences of the donor and acceptor vectors must be able to
10 recombine intermolecularly (e.g., L1 with a compatible L1, and L2 with a compatible L2) with one another to allow DNA exchange between the donor and acceptor vectors. In order to ensure intermolecular exchange between compatible lox sequences, the lox sequences are generally oriented in the same direction.

Incompatibility between locks sequences can be achieved, for example, by way
15 of mutating or modifying (e.g., by nucleotide addition, deletion or substitution) one of two identical lox sequences, preferably in their spacer sequences, so that the sequences differ. Testing to determine which mutations confer incompatibility can be performed using standard mutation assays which test for the ability of the mutated and non-mutated lox sequences to recombine.

20 In a preferred embodiment, one of the two incompatible lox sequences is the Lox P1 sequence of the Cre/lox system of bacteriophage P1 (Hoess et al. (1990) "Nucleic Acids and Molecular Biology," Vol 4, p. 99) having the sequence shown in SEQ ID NO: 1. The Lox P1 sequence is a 34 base pair sequence which can be isolated from bacteriophage P1 by methods known in the art (see e.g., Hoess et al. (1982) *PNAS*
25 79:3398). The Lox P1 sequence consists of two 13 base pair inverted repeats separated by an eight base spacer sequence. Lox P1 sites can also be isolated from plasmids available from the ATCC (e.g., ATCC 53254 and 20773). Other suitable lox sequences include the Lox B, Lox L, and Lox R sequences isolatable from *E. coli* (Hoess et al. (1982), *supra*). Lox sequences can also be chemically synthesized using known
30 techniques, such as those described in the Examples below.

Accordingly, the other incompatible lox sequence can be a mutated form of the LoxP1 sequence, for example, having a point mutation in the eight nucleotide spacer sequence. In one embodiment, the point mutation is substitution of A for G at position 7 of the eight base spacer sequence of the wild type Lox P1 sequence, referred to herein as the Lox511 sequence (SEQ ID NO: 2). Accordingly, in one embodiment, the two incompatible lox sequences of the invention have the following sequences:

SPACER

10	Lox P1	ATAACTTCGTATA ATGTAT <u>GC</u> TATACGAAGTTAT
	Lox 511	ATAACTTCGTATA ATGTAT <u>AC</u> TATACGAAGTTAT

Alternatively, the donor and acceptor DNAs (e.g., vectors) contain two identical but inverted lox sequences so that intramolecular recombination between the inverted lox sequences results in an inversion of the intervening DNA, and intermolecular recombination between identical lox sequences results in exchange of donor and acceptor DNA (see Figure 4). Moreover, because the lox sequences involved are all substantially identical, the frequency of recombination among them is the same. These features provide highly (up to 100%) efficient and stable recombinase-mediated DNA integration. In fact, in one embodiment, cells which have taken up the largest amount of donor DNA, as measured by e.g., a transient, non-integrated marker gene (e.g., green fluorescence protein), can be selected for and, among these cells, DNA exchange will be high enough that no selectable marker need be used. In these embodiments, as in those described in the preceding paragraphs, suitable inverted lox sequences include but are not limited to the LoxP1 (SEQ ID NO:1), Lox511 (SEQ ID NO:2) and inversions thereof.

Intermolecular recombination between compatible or inverted lox sequences in the donor and acceptor vectors is catalyzed by a recombinase, such as Cre or another member of the Int family of recombinases (Argos et al. (1986) *EMBO J.* 5: 433) which have been shown to perform efficient recombination at lox sequences in both bacteria and in eukaryotic cells (Sauer et al. (1993) *Methods in Enzymology* 225: 890-900). The recombinase can be introduced into the cell along with the donor and acceptor vectors in

the form of a protein or as an expressible gene encoding the protein (e.g., the Cre gene described by Sauer, B. et al. (1988) *Proc. Natl. Acad. Sci. USA* 85:5166-5170). The recombinase or recombinase gene can be introduced or transfected into the host cell before, simultaneously, or following introduction of the donor and acceptor vectors.

5 In one embodiment, the recombinase gene (e.g., Cre) is contained in an expression vector which is co-transfected with the donor vector following introduction and integration of the acceptor vector into the host cell. In another embodiment, the recombinase gene is contained within either the acceptor vector or the donor vector. As with the donor vector, the recombinase gene can be introduced into the host cell either *in vivo* or *in vitro* using known techniques, such as CaPO₄ precipitation, electroporation, 10 cationic lipofection, use of artificial viral envelopes, microinjection (e.g., into fertilized eggs), direct injection (e.g., intravenous, intraperitoneal or intramuscular). The vector can also be targeted for delivery to selected cells (see e.g., US 5,166,320).

The DNA which is transferred from the donor to the acceptor vector by way of 15 the site-specific recombination method of the invention can be any DNA desired for stable integration into a host cell genome. For example, the gene can be any transgene useful, for example, in gene therapy or for diagnostic purposes. The gene can encode a desired therapeutic protein, such as α , β or δ globin, blood coagulation factors (e.g., Factors VIII and IX) gene, cell surface receptors and other desirable proteins, for 20 example, to correct inherited deficiencies of these proteins in an individual.

Accordingly, in one embodiment, the invention can be used *in vivo* to create transgenic mammals and/or plants. For example, animals of human disease can be generated, particularly if multiple genes have to be expressed at well regulated levels. Transgenic animals and/or plants can be created using known techniques, such as 25 microinjection or other methods of introducing DNA into fertilized eggs or cells (e.g., embryonic stem cells), which contain target lox sequences (e.g., inverted or incompatible) at chromosomal locations that are not subject to position effects or to desired position effects for directing expression of a gene of interest. This allows for the generation of animals and/or plants with, for example, higher resistance to disease or 30 improved physical/functional characteristics.

In general, there are two broad types of applications for the invention depending on whether the Lox sites are integrated in the genome at random sites or at known chromosomal locations (e.g., via homologous recombination). Lox sites at random locations can be used to generate reference loci, such as loci with characteristic position-effects. For instance, the invention can be used to site-specifically integrate particular genes into reference loci favorable for strong expression in a particular tissue, or loci that confer a pan-cellular or a variegated pattern of expression. Moreover, site-specifically integrated donor DNA at known genomic locations can facilitate the systematic characterization of endogenous cis-regulatory elements and the genetic dissection of protein function via the generation of mutated versions of endogenous gene products.

Accordingly, in other embodiments, the invention provides a means for (1) the discovery and analysis of cis-acting DNA regulatory elements controlling expression and position-effects 2) the production of polypeptides of biological value by insertion of transgenes at reference loci at which expression is high or optimally inducible, (3) the development of designer proteins by insertion of multiple versions of a gene at the same locus in order to test variants in a context in which they are produced in the same amount and have the same tissue distribution, (4) the creation of animal models of human diseases (particularly those in which multiple genes must be expressed at well regulated levels), and (5) the creation of improved agricultural organisms.

Classical gene transfer technology often involves cumbersome screening procedures to identify clones or animals with appropriate levels and developmental patterns of expression. These first generation technologies are sufficient to perform simple genetic manipulations, however, they are inadequate for the implementation of complex genetic strategies to modify plants and animals by introduction of large numbers of carefully regulated transgenes to achieve a desired phenotype. For such complex genetic engineering purposes the development of rules for the creation of artificial genetic loci (AGL) is required.

The compositions and methods of the present invention provide a means for creating an AGL by allowing for the production of an array of genes and regulatory elements integrated at reference genomic sites whose chromatin structure and influence on gene

expression can be controlled at will by insertion of known cis-acting elements. Thus, transgenes making up an AGL can be expressed in a tissue and in a developmental stage-specific manner at levels that are totally predictable. The present invention facilitate the development of AGLs by both permitting the careful characterization of the site with different cis-regulatory elements and mediating the efficient and convenient exchange of multiple transgenes into the AGL. Accordingly, the methods and compositions of the invention can be used for a variety of therapeutic and diagnostic applications which require stable and efficient integration of transgene sequences into genomic DNA of cells. The methods and compositions can be used to transform a wide variety of eukaryotic cells (e.g., mammalian) cells and provide the advantage of high efficiency DNA transfer.

EQUIVALENTS

Those skilled in the art will recognize, or be able to ascertain using no more than routine experimentation, many equivalents to the specific embodiments of the invention described herein. Such equivalents are intended to be encompassed by the following claims. The entire contents of all references, published patent applications and issued patents cited throughout this application are hereby incorporated by reference.

EXAMPLES

EXAMPLE 1 - CRE/LOX MEDIATED GENE INTEGRATION USING INCOMPATIBLE MUTATED LOX SEQUENCES

DNA Construction and Cell Culture

DNA vectors were made using standard techniques (Sambrook, J. et al. (1989) *Molecular Cloning: A Laboratory Manual - 2nd ed.* Cold Spring Harbor Laboratory Press, Cold Spring Harbor, New York, USA). Oligonucleotides were synthesized by Research Genetics, Inc. Accuracy of DNA construction was verified by sequencing. LXSN retroviral vector (Miller, A.D. et al. (1989) *Biotechniques* 7:980-990) was provided by D. Miller (Fred Hutchinson Cancer Research Center, Seattle), Hygromycin B (Lupton, S.D. et al. (1991) *Mol. Cell. Biol.* 11:3374-3378) phosphotransferase gene by

D. Housman (MIT, Cambridge), Herpes Simplex virus thymidine kinase (HSV-TK) gene (Lupton, *supra*) by M. R. Capecchi (Salt Lake City, Utah), U19 SV40T mutant gene (Renfranz, P.J. et al. (1991) *Cell* 66:713-729) by R. D. McKay (MIT, Cambridge) and G. Almazan (McGill University, Montreal), Cre recombinase gene (Sauer, B. et al. 5 (1988) *Proc. Natl. Acad. Sci. USA* 85:5166-5170) by D. W. Ow (UC Berkeley, Albany), CD24 (Pawlink, R. et al. (1994) *Blood* 84: 2868-2877). MSCV (murine stem cell virus) retroviral vector (Hawley, P.G. et al. (1994) *Gene Therapy* 1:136-138), pBabe retroviral vector (Morgenstern, J. P. et al. (1990) *Nucl. Acids Res.* 18:3587-3596) by R. Weinberg (MIT, Cambridge), pcDNA1 by Invitrogen Corp. and pOPRSVICAT by Stratagene, 10 Inc. NIH3T3 cells were obtained from the ATCC, BOSC23 cells (Pear, W. S., et al. ((1993) *Proc. Natl. Acad. Sci., USA* 90:8392-8396) W. Pear and D. Baltimore (Rockefeller University, New York).

NIH3T3 cells were grown at 37°C with 5% CO₂/95% air in DMEM supplemented with 10% heat inactivated calf serum (CS), 4.5 mg/ml glucose, 2 mM 15 glutamine, 100 IU/ml penicillin and 100 µg/ml streptomycin. For BOSC23 cells, CS was replaced by 10% heat inactivated fetal calf serum (FCS).

Cell Infection, Transfection and Selection

The packaging cell line, BOSC23, was grown as described (Pear, *supra*, Danos, 20 O. et al. (1988) *Proc. Natl. Acad. Sci. USA* 85:6460-6464). Plasmid DNA's were prepared by the Qiagen procedure (Qiagen, Inc.) and transfected in BOSC23 cells using a calcium phosphate procedure (5prime:3prime, Inc.). Viral supernatants from producers were harvested and filtered as described (Pear, *supra*, Danos, *supra*). All infections were carried out in the presence of 8 µg/ml Polybrene (Sigma). Viral 25 supernatants from BOSC23 were used to generate stable viral producers. Virus titers were estimated by infection and selection of NIH3T3 cells using standard calculations previously described (Pear, *supra*, Danos, *supra*). Detection of helper viruses was performed by a β-galactosidase mobilization assay as described (Pear, *supra*, Danos, *supra*). Selection was applied two days after infections.

Standard concentrations (1 X) of selection agents were 320 µg/ml for Hygromycin B (Calbiochem). Packaging NIH3T3 cells were selected with 1X, MDHF with 1/2 X and BSMC with 2 X concentrations.

5 Site Specific DNA Integration Using Incompatible Mutated Lox Sequences

To study the efficiency of gene integration using the Cre/lox mediated gene transfer system described herein, the following protocol was performed.

10 Acceptor vectors were constructed using the MSCV retroviral vector. The vectors contained in order: the left MSCV LTR (containing promoter), followed by a lox L1 sequence, followed by a hygromycin-TK fusion gene (as a selectable marker), followed by a lox L2 sequence, followed by the right MSCV LTR (see Figure 2). The retrovirus LTR was used as a the promoter for the hygromycin-TK fusion gene. Similar constructs were made using other selection markers such as neomycin.

The L1 and L2 lox sequences of the acceptor vector had the nucleotide sequences shown below (corresponding to SEQ ID NO: 1 and SEQ ID NO: 2). L1 is the wild type LoxP1 sequence (SEQ ID NO: 1) from bacteriophage P1 (Abremski et al. (1983) *Cell* 32: 1301-1311). L2 is a mutated form of the wild type LoxP1 sequence, referred to as Lox511, having a point substitution of A for G at position 7 of the eight nucleotide spacer region (Waterhouse et al. (1993) *Nucleic Acids Res.* 21(9):2265-2266).

20

SPACER

L1	ATAACTTCGTATA ATGTATGC TATACGAAGTTAT
L2	ATAACTTCGTATA ATGTATAC TATACGAAGTTAT

25 Following construction of the acceptor vector ("L1-hygromycin-TK-L2 construct"), BOSC23 cells (ecotropic packaging cells) were transiently transfected with acceptor vector using a calcium phosphate procedure (5prime:3prime, Inc.). Viral supernatants from producers were harvested and filtered as described (Pear, *supra*, Danos, *supra*). All infections were carried out in the presence of 8 µg/ml Polybrene
30 (Sigma) Viral supernatant containing high-titer ($>10^5$ pfu/ml) retroviral vector was then

used to infect host NIH 3T3 cells using the same procedures. After 48 hours in culture, the infected 3T3 cells were selected with hygromycin.

Donor vectors were constructed using pUC 19 plasmid (Yanish-Perron et al. (1985) *Gene* 33:103-119) as the backbone. The vectors contained in order: the L1 lox sequence, followed by a promoterless neomycin gene, followed by the L2 lox sequence (see Figure 2). Similar donor vectors were made using hygromycin-TK, CD24, and B-globin genes in place of the neomycin gene. Control donor vectors were constructed using a neomycin gene with the PGK (phosphoglycerol kinase) promoter, PGK-neomycin.

10 Various concentrations of donor vector containing neomycin gene were co-electroporated, along with an expression vector containing the Cre recombinase gene, into the infected 3T3 cells. The concentrations of donor vector ranged from 10 µg to 200 µg. After 48 hours in culture, transformed cells were selected with neomycin. Concentrations 100 µg or more of donor vector resulted in a 10-30% integration efficiency (as measured by transfer of neomycin gene for hygromycin gene).

Different ratios of donor vector and Cre expression vector, ranging from 20:1 to 1:1 were co-electroporated into the infected 3T3 cells. All ratios resulted in the transfer of the neomycin for the hygromycin. However, a ratio of 3:1 (donor:Cre) resulted in the highest integration efficiency.

20 The following table provides the results of neomycin gene integration using various donor and Cre expression vectors (see Figure 3) at a concentration of 100 µg of donor vector (DNA) at a ratio of 3 parts donor vector to 1 part Cre expression vector. Experiments E#1-4 were performed as negative controls. E#5 was the positive control.

- 21 -

	Cells Electroporated	Constructs Used	# Colonies (out of 10^7 cells)
5	E#1 3T3	lox 1-PGKNeo-lox 2 & control Cre expression vector	530
	E#2 3T3	lox 1-Neo-lox 2 & control Cre expression vector	10
10	E#3 3T3	lox 1-Neo-lox 2 & Cre expression vector	2
	E#4 3T3 containing lox A-hygro-TK-lox B	lox 1-Neo-lox 2 & control Cre expression vector	21
15	E#5 same	lox 1-Neo-lox 2 & Cre expression vector	confluent ($>10^5$)

E#1 used the control donor vector (see Figure 2), lox 1-PGKNeo-lox 2 (containing the neomycin gene and a promoter) along with a control Cre expression vector (see Figure 3) (in which the sequence encoding Cre had been deleted and replaced by a gene encoding CAT). Host cells did not contain integrated acceptor vector. Therefore, E#1 demonstrated the amount of neomycin resistance conferred by random integration of the L1-PGKNeo-L2 vector capable of expressing the neomycin gene. As expected, the conferred neomycin resistance was in the range of efficiency of integration obtained by electroporation (e.g., about 0.1% efficiency).

E#2 used donor vector (promoterless) with a control Cre expression vector. Host cells did not contain integrated acceptor vector. Therefore, E#2 demonstrated the resistance conferred in the absence of acceptor vector or Cre recombinase (i.e., in the absence of efficient recombination and gene transfer). As expected, this was very low.

E#3 used donor vector (promoterless) with a functional Cre expression vector. Host cells did not contain integrated acceptor vector. Therefore, E#3 demonstrated the resistance conferred in the absence of acceptor vector (i.e., in the absence of efficient recombination and gene transfer), but in the presence of Cre recombinase. As expected, this was very low.

E#4 used donor vector (promoterless) with a control Cre expression vector (no Cre expression). Host cells contained integrated acceptor vector (L1-hygro-TK-L2). Therefore, E#4 demonstrated the gene transfer efficiency from donor vector to acceptor vector in the absence of Cre). As expected, this was very low.

E#5 used donor vector (promoterless) with a functional Cre expression vector. Host cells contained integrated acceptor vector (L1-hygro-TK-L2). Therefore, E#5 demonstrated the gene transfer efficiency from donor vector to acceptor vector in the presence of Cre. As shown in the table above, the host cells became confluent, demonstrating a greater than 1000 fold increase in gene transfer efficiency and stability.

Conclusion:

The foregoing studies and results demonstrate that the retroviral Cre/lox mediated gene transfer system of the present invention can be used for highly efficient and stable integration of transgenes into chromosomal DNA of mammalian cells.

EXAMPLE 2 - CRE/LOX MEDIATED GENE INTEGRATION USING

INVERTED LOX SEQUENCES

Donor and Acceptor Plasmids

Plasmid constructions were performed as described above in Example 1 using standard procedures. The CMV-HYTK cassette was derived from plasmid pTgCmwhytk (Immunex, Seattle, Wa). The CMV-EGFP and GFPuv cassettes were derived from plasmids pEGFP-N1 and pGFPuv (Clontech, Palo Alto, CA). The β -GFP cassette was created by replacing the CMV promoter in plasmid pEGFP-N1 by the β -globin

promoter (fragment -374 to +44 relative the cap site). Cassette HS234- β GFP was produced by linking fragment HS234 (Bouhassira et al. (1997) *Blood* 90:3332-3344) to the β -GFP cassette. Cassette HS3- β -globZ was from pCAS3 (O'Gorman et al. 1991) *Science* 251:1351-1355. A series of plasmids containing a cassette flanked by Lox sites, L1 (SEQ ID NO:1) and L2 (SEQ ID NO:2), in various orientations were created. Each plasmid was designated by the Lox site 5' of the cassette, the name of the cassette and the Lox site 3' of the cassette. Inverted L1 Lox site were designated as 1L.

The following plasmids were used in the studies described below: pL1-HYTK-L2, pL1-HYTK-1L, pL1-CMVEGFP-L2, pL1-Gfpuv-L2, pL2-GFPuv-L2, pL1-HS234 β GFP-1L, pL1-HS3 β globZ-L2EGFP (contains EGFP 3' of the L2 Lox site).

Cre Expression Plasmid (mammalian):

Plasmid pBS 185 (CMV-CRE) was obtained from Clontech (Palo Alto, CA). pSSR73 (RSV-CRE) was a gift of Dr. P. Leboulch (Harvard University, Cambridge MA), pMC-Cre was a gift of Klaus Rajewsky (Koln, Germany) (Gu et al. (1993) *Cell* 73:1155-1164).

Cre Expression Plasmid (bacterial):

A PCR generated fragment containing the coding sequence of the CRE recombinase was cloned into the NdeI/andXhoI sites of plasmid pET23c (Novagen, Madison, WI) resulting in a plasmid (pET23c-CRE) that expresses a His-tagged CRE protein in induced BL21 bacteria.

CRE Purification:

BL21 (DE3) Lys E bacteria (Novagen, Madison, WI) containing plasmid pET23c-CRE were induced with 2 mM IPTG for 2 hours, sonicated in basic buffer (50mM NaPi pH 8, 200mM NaCl) plus 0.05% Tween 20 and 1% lysozyme, centrifuged at 4000g for 10 min at 4°C, and the supernatant was loaded on a Talon metal affinity Resin column (Clontech, Palo Alto, CA). The column was washed successively with 20 ml of basic buffer containing 0, 5 and 10 mM imidazol, and the CRE protein was then

eluted with 5 ml of basic buffer containing 100mM imidazol. In vitro CRE recombination were performed as in ¹⁸.

Cell Culture and Electroporation:

5 Culture and DNA recombination reactions in MEL cells were performed as in (Bouhassira et al. (1997) *Blood* 90:3332-3344). AK-7 ES cells were cultured and selected on SNL feeder cells as described (Soriano P. (1997) *Development* 124:2691-2700). DNA recombination was performed by electroporating (250V, 500uF, Biorad gene pulser, Biorad, Hercules, CA) 3 x 10⁶ cells with 200 µg of L1GFP1L and 30 µg of
10 MC-Cre. Selection with 3 µM gancyclovir was applied 5 days after the transfection.

GFP Expression Studies:

5 x 10⁵ cells were rinsed once in PBS and resuspended in Hank's solution containing 5% Fetal Calf Serum and 2 µM propidium iodide, and analyzed on a
15 FACSCAN flow cytometer (Becton Dickinson, CA).

Site Specific DNA Integration Using Inverted Lox Sequences

The following studies were performed to determine whether the efficiency of site-specific DNA integration could be increased using two inverted lox L1 (SEQ ID
20 NO:1) sequences, in place of the lox L1 and L2 sequences used in Example 1. Thus, the L2 site was replaced with an inverted L1 site (termed 1L). As shown in Fig. 4 (panel C), intra-chromosomal recombination between two inverted Lox sites leads to inversion of the HYTK negatively selectable marker previously integrated into the genome, rather than to its excision, and therefore does not remove the sensitivity to negative selection
25 by gancyclovir. In this system, the donor cassette integrates in one orientation in half of the clones and in the reverse orientation in the other half.

A plasmid (pL1-HYTK-1L) containing the HYTK gene flanked by two inverted but identical Lox L1 (SEQ ID NO:1) sites was created, and MEL cell lines with single integrated copies of this plasmid were produced as described above. Three clones,
30 termed RL4, RL5 and RL6 were selected for further studies. Exchange reactions were performed as above using a plasmid in which inverted L1 Lox sites flank a GFP reporter

(pL1-HS234GFP-1L). After the transfection, 12 Gan^R clones per cell line were picked and analyzed by Southern blots. At loci RL4 and RL5, all 12 Gan^R clones tested had undergone an exchange of the HYTK cassette for the GFP cassette (Fig. 5 (panel A)). At RL6, 11 out of 12 clones tested had undergone an exchange of the HYTK cassette for the GFP cassette. As expected, about half of the clones were in reverse orientation. These results demonstrate that DNA cassette exchange with inverted lox sites can be efficiently selected using only negative selection in MEL cells.

The same experiments were then repeated in an embryonic stem cell line: Two single copy L1-HYTK-1L tagged-loci were created and tested for cassette exchange using plasmid pL1-CMVEGFP-1L. At the first locus tested, about 50% of the 60 Gan^R clones tested by Southern blot had undergone an exchange (Fig. 5 (panel B)). At the second locus the frequency of exchange was about 10%. As in MEL cells, clones in both orientations were obtained in approximately equal proportion.

15 **EXAMPLE 3 - CRE/LOX MEDIATED GENE INTEGRATION USING NO SELECTABLE MARKER**

The materials and methods used in the studies described below were as described in Examples 1 and 2 above.

Typically, site specific DNA recombination systems employ an active negative selectable marker at the integration site prior to exchange. Since, in many situations, it would be advantageous to perform recombination at loci that do not contain any active genes at all, the following system was designed which does not require the presence of an active gene before or after the recombination (e.g., cassette exchange).

When no selection is applied, about 1% of the MEL cells that survive the transfection undergo recombination. To increase this frequency, a GFP reporter gene was placed in the exchange plasmid, outside the exchange cassette, and cells that expressed the highest levels of GFP were selected for to increase the proportion of selected cells which had undergone exchange. This is because the sorted cells would be the cells that have taken up the highest amount of the exchange plasmid during the transfection (Figure 4 (panel D)).

Specifically, a plasmid containing a GFP reporter gene located outside of a LacZ exchange cassette (pL1-HS3 β globZ-L2-GFP) was created and co-transfected with a CRE expression plasmid in two lines of MEL cells (RL1 and RL3) containing pre-integrated target L1 and L2 Lox sites flanking the HYTK gene. Forty-eight hours post-transfection, the cells expressing the highest amount of GFP (0.1 percentile) were sorted individually into 96 well plates, expanded without applying any selection and tested for DNA cassette exchange. Seventy-four clones were obtained for the RL1 line and 32 for the RL3 line. The clones that had lost the HYTK gene were identified by culture in hygromycin, and the Hyg^S clones were then tested by Southern blots: At locus RL1, 4% (3/74) of the clones obtained had an exchange and 12% had lost the locus without exchange, probably because of a CRE-mediated excision. At locus RL3, 16% (5/32) of the clones had an exchange and 12% had lost the locus without exchange. This demonstrates that exchange with no expressed selectable marker in the genome can be performed at relatively high frequency by simply sorting the cells that have been transfected with a large number of exchange plasmids.

EXAMPLE 4 - SITE-SPECIFIC INTEGRATION GREATLY IMPROVES REPRODUCIBILITY OF EXPRESSION

The materials and methods used in the studies described below were as described in Examples 1 and 2 above.

To determine whether site-specific chromosomal integration leads to expression that is more reproducible and predictable than expression of the same cassette, randomly integrated. Mel cell clones with integration of the CMV-EGFP cassette at random sites were generated by transfection of plasmid pEGFP-N1, and compared with clones containing the CMV-EGFP cassette integrated by site-specific Cre/lox-mediated exchange at locus RL4.

Southern blots revealed that most of the random clones contained multiple integrated copies (Figure 5 (panel D)) and that, as expected, the clones that had undergone site-specific exchange had integration in both possible orientations. Nine random clones and nine site-specific clones with integration in one of the two possible orientations were analyzed for expression by flow cytometry (fig. 3). Mean levels of

green fluorescence for the random clones varied from 5.3 (which is similar to untransfected controls) to 357.5. The average was 94.1 ± 103.5 . Mean levels of fluorescence for the site-specific clones varied from 100.1 to 147.4. The average was 127.8 ± 11.9 . Although the average expression of the two types of clones were similar, the 10-fold difference between their standard deviations clearly demonstrates that expression levels of clones integrated at the same locus by site-specific recombination are much less variable than those of randomly integrated clones. Similar results were obtained at the 5 other targeted, site-specific loci that were tested.

The highly reproducible expression observed when site-specific clones were tested in the absence of any selective pressure at the locus suggests that controlling the integration site and therefore eliminating position-effects permits meaningful comparisons of multiple transgenes individually integrated at a particular genomic site.

Conclusion

Overall, the studies described above demonstrate that highly efficient methods of Cre/lox-mediated recombination, such as those described in the studies above, can be used to achieve site-specific recombination with no incoming selectable marker and in a defined orientation. These methods also can be used to achieve multiplex integration of transgenes (Sauer B. (1996) *Nucleic Acids Res.* **24**:4608-4613) sequentially at the same site or at multiple independent sites using multiple pairs of lox sites, either as heterospecific pairs or inverted as described herein. These methods also can be used in combination with FLP (Seibler et al. (1998) *Biochem.* **37**:6229-6234)-mediated DNA exchange systems.

Prior to the present invention, expression of experimental constructs in mammalian cells or transgenic animals was difficult to control because it is markedly influenced by position effects. This limited both the analysis of cis-DNA regulatory elements for transcription and replication and the physiological analysis of proteins expressed from transgenes. The novel compositions and methods of the present invention solve these problems by permitting the exchange of DNA (e.g., containing a negative selectable marker pre-localized on the chromosome with a transgene via a Cre-mediated double recombination between inverted or mutated, incompatible Lox sites.

Integration efficiency of close to 100% of negatively selected cells can be achieved. In addition, the present invention provides a way of achieving site-specific recombination within genomic DNA with no selection at all, except for cells that have taken up plasmid transiently.

- 5 The foregoing studies demonstrate that integration of a transgene at a given genomic site leads to reproducible expression. Therefore, the compositions (e.g., DNA vectors) and methods of the present invention can be used to develop artificial genetic loci that impart specific and reproducible regulation of transgenes in higher eukaryotes. This should facilitate the analysis of cis-regulatory DNA elements governing expression
- 10 and position effects, improve control over the physiological effects of transgenes, and accelerate the development of animal models for complex human diseases.

What is claimed is:

1. A method of achieving site-specific recombination comprising, in any order, the steps of:
 - 5 (a) providing a donor DNA comprising two inverted lox sequences;
 - (b) providing an acceptor DNA comprising the same two inverted lox sequences contained in the donor DNA; and
 - (c) contacting the donor and acceptor DNA with a recombinase.
- 10 2. The method of claim 1 further comprising the step of integrating the acceptor DNA into the genome of a cell.
3. The method of claim 1 wherein the donor DNA is present in excess of the acceptor DNA.
- 15 4. The method of claim 1 wherein the recombinase is selected from the group consisting of Cre recombinase.
5. The method of claim 1 wherein the recombinase is selected from the
20 group consisting of Flp recombinase
6. The method of claim 4 wherein the inverted lox sequences comprise a nucleotide sequence selected from the group consisting of SEQ ID NO:1, SEQ ID NO:2 and inversions thereof.
- 25 7. The method of claim 2 wherein the acceptor DNA is integrated into the genome using a retroviral vector.
8. The method of claim 2 wherein the acceptor DNA is integrated into the
30 genome by homologous recombination.

9. A method of achieving site-specific recombination in a selected cell comprising, in any order, the steps of:

(a) introducing into the cell an acceptor DNA which integrates into the genome of the cell, the acceptor DNA comprising two inverted lox sequences;

5 (b) introducing into the cell a donor DNA comprising the same two inverted lox sequences contained in the acceptor DNA; and

(c) introducing into the cell a recombinase which catalyzes recombination between the lox sequences.

10 10. The method of claim 9 wherein the acceptor DNA further comprises a negatively selectable marker flanked by the two inverted lox sequences.

11. The method of claim 10 further comprising the step of negatively selecting cells which have undergone site-specific recombination.

15

12. The method of claim 9 wherein the donor DNA introduced into the cell is present in excess of the acceptor DNA.

13. The method of claim 9 wherein the recombinase is introduced into the cell by transfecting the cell with a gene encoding the recombinase.

20

14. The method of claim 9 wherein the recombinase is introduced into the cell in the form of a protein.

25 15. The method of claim 9 wherein the inverted lox sequences comprise a nucleotide sequence selected from the group consisting of SEQ ID NO:1, SEQ ID NO:2 and inversions thereof, and the recombinase is Cre.

16. The method of claim 9 wherein the donor DNA further comprises a transgene encoding a therapeutic protein.

30

17. The method of claim 16 wherein the protein is β -globin.
18. The method of claim 9 wherein the donor DNA further comprise a positively selectable marker.
- 5 19. The method of claim 9 wherein the cell is a mammalian cell.
20. The method of claim 9 wherein the donor DNA, the acceptor DNA or both the donor DNA and the acceptor DNA are introduced into the cell by
10 microinjection into a fertilized egg.
21. The method of claim 9 wherein the donor DNA, the acceptor DNA or both the donor DNA and the acceptor DNA are introduced by microinjection into an embryonic stem cell.
- 15 22. A DNA vector comprising two inverted lox sequences.
23. The vector of claim 22 wherein the inverted lox sequences comprise a nucleotide sequence selected from the group consisting of SEQ ID NO:1, SEQ ID NO:2
20 and inversions thereof.
24. The vector of claim 22 further comprising a selectable marker.
25. The vector of claim 22 which is a retroviral vector or an adeno-associated
25 viral vector.
26. A gene transfer system comprising:
(a) acceptor DNA which integrates into the genome of a host cell, the acceptor DNA comprising two inverted lox sequences; and
30 (b) donor DNA comprising a transgene flanked by the same two inverted lox sequences contained in the donor DNA.

27. The gene transfer system of claim 26 further comprising a recombinase which catalyzes recombination between the inverted lox sequences, or DNA encoding the recombinase.

5

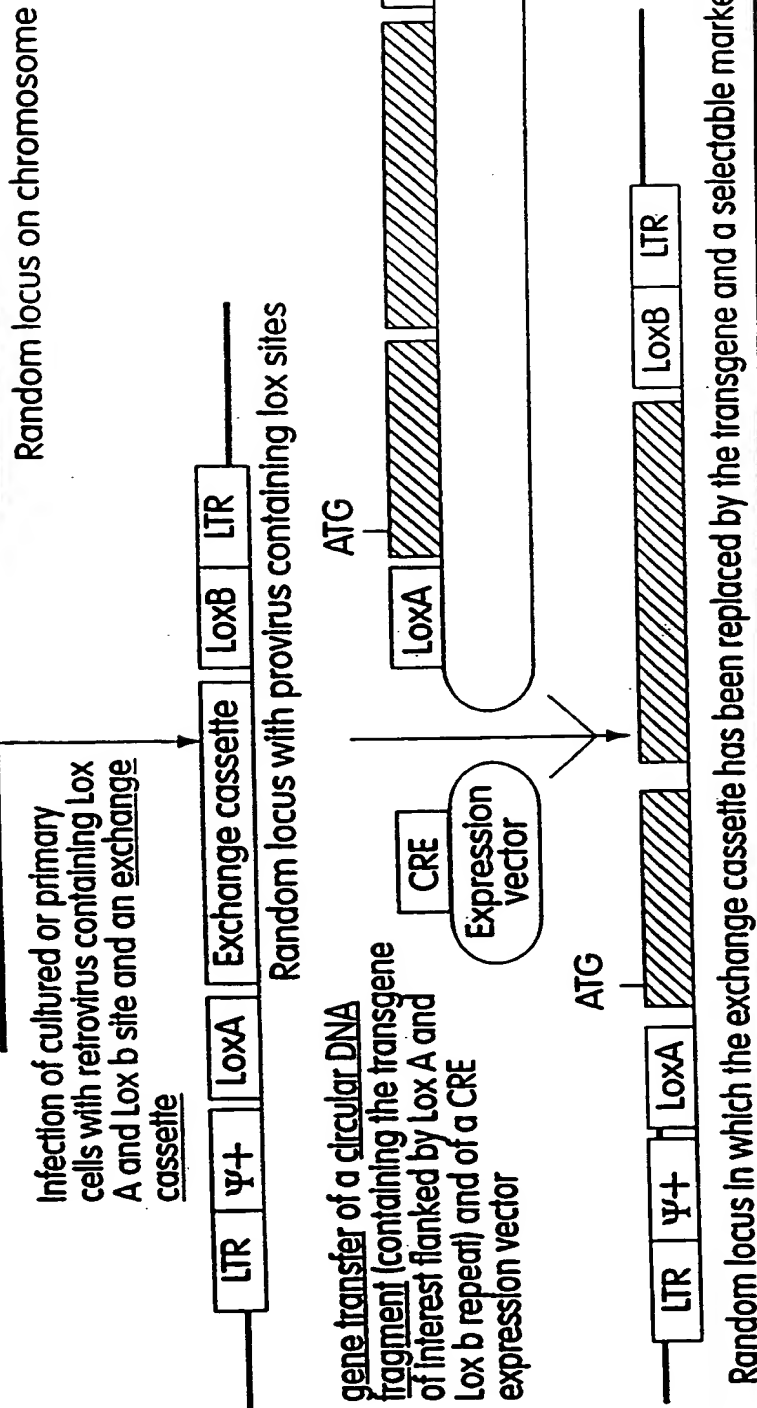
28. The gene transfer system of claim 26 wherein the acceptor DNA further comprises a negatively selectable marker.

29. The gene transfer system of claim 26 wherein the donor DNA further
10 comprises a positively selectable marker.

30. The gene transfer system of claim 26 wherein the inverted lox sequences comprise a nucleotide sequence selected from the group consisting of SEQ ID NO:1, SEQ ID NO:2 and inversions thereof.

15

RETROVIRUS-CRE-LOX-INTEGRATION



In this diagram the Cre enzyme is cotransferred as a plasmid. It could also be provided as a protein or be in the exchange cassette

CRE
Expression vector

— = Chromosome

LoxA AND LoxB = 2 mutually incompatible Lox sites

Exchange cassettes: Can be an antibiotic resistance gene or fusion gene (such as Hyg), a gene that can be used to identify the infected cells by FACS or any other means, or an irrelevant DNA fragment

Gene transfer: can be by electroporation, lipofection, or any other means of introducing DNA into the cells

LEGENDS AND DEFINITIONS

LTR = Long Terminal Repeat

Psi+ = packaging signal

ATG

Antibiotic resistance gene or gene selectable by any other means

Antibiotic resistance

gene or gene selectable by any other means

Circular DNA fragment: Plasmid, Cosmid or other similar vector as well as DNA fragment circularized by enzymatic reactions

Fig. 1

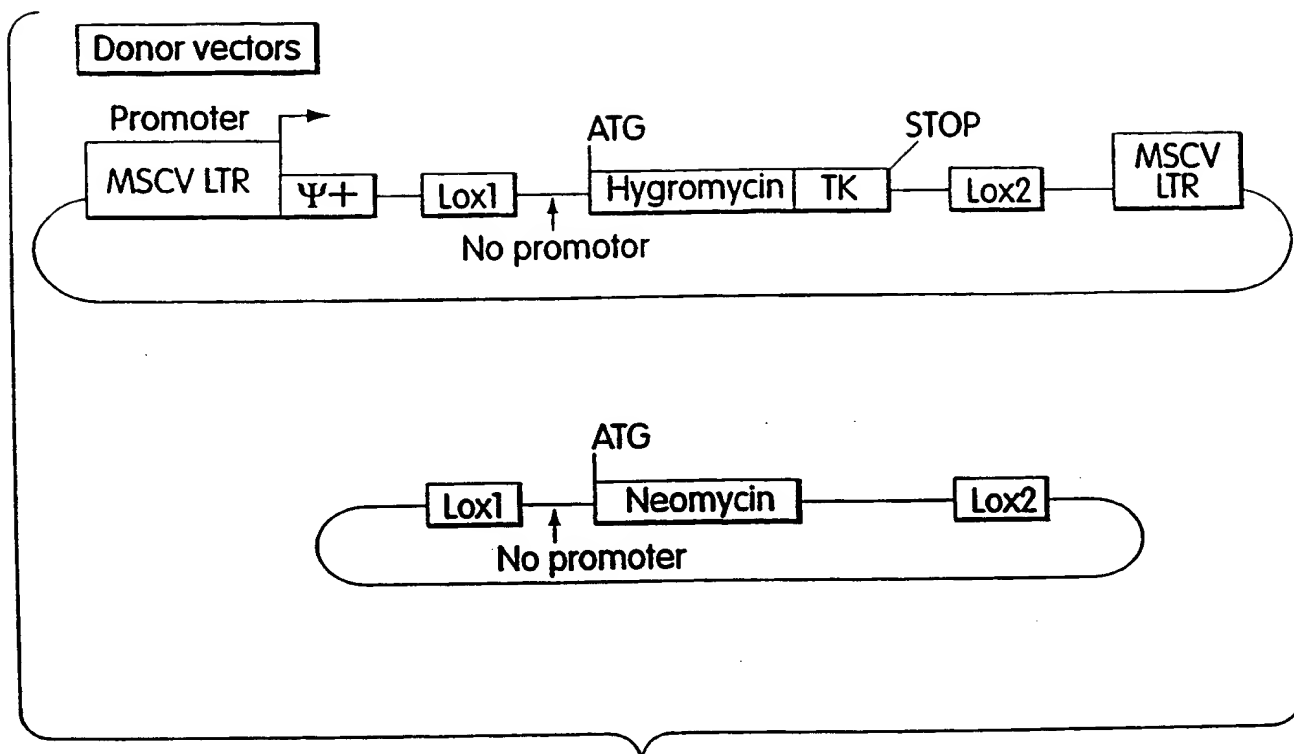


Fig. 2

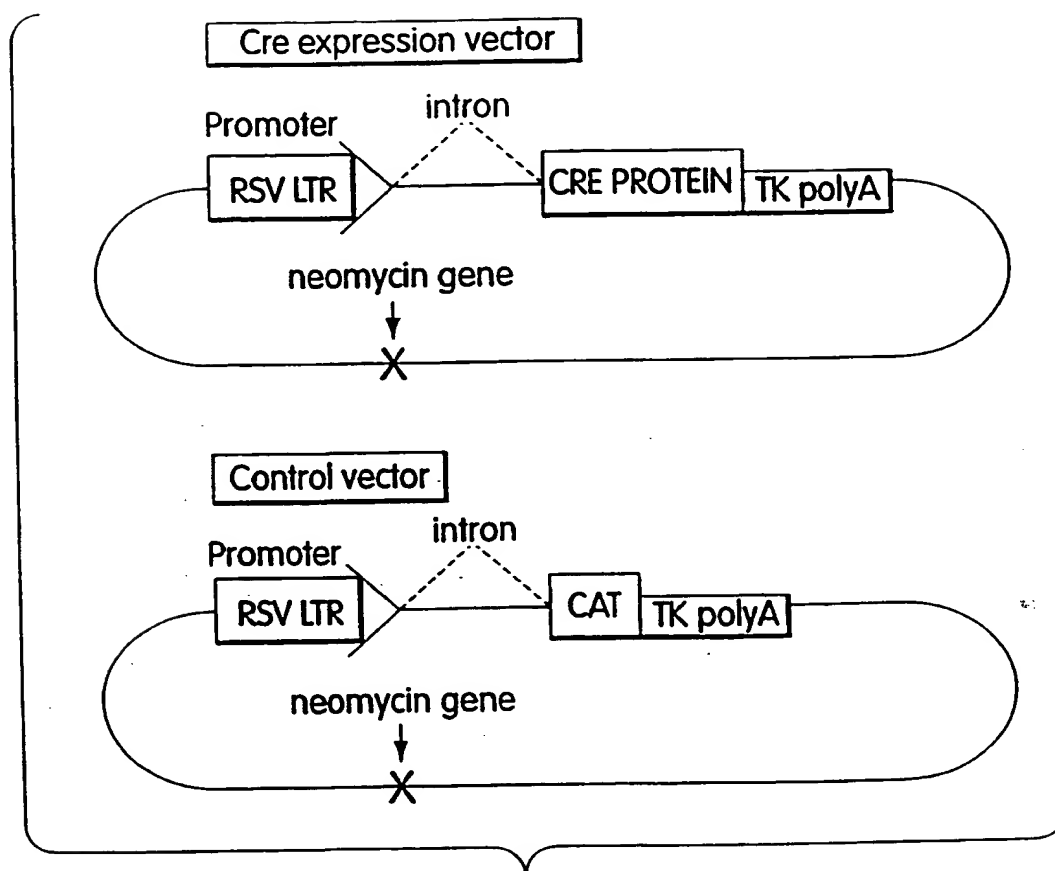


Fig. 3

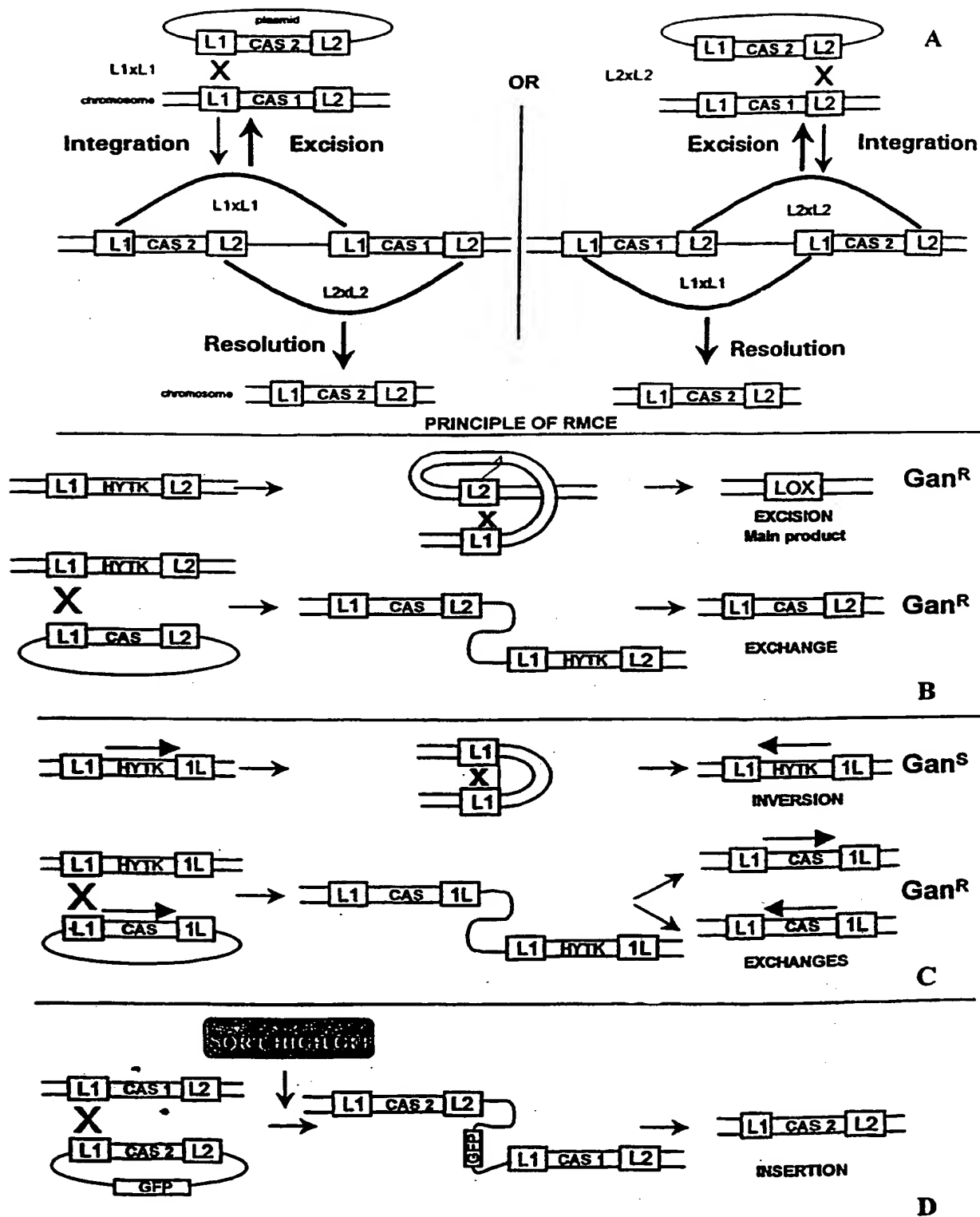


FIGURE 4

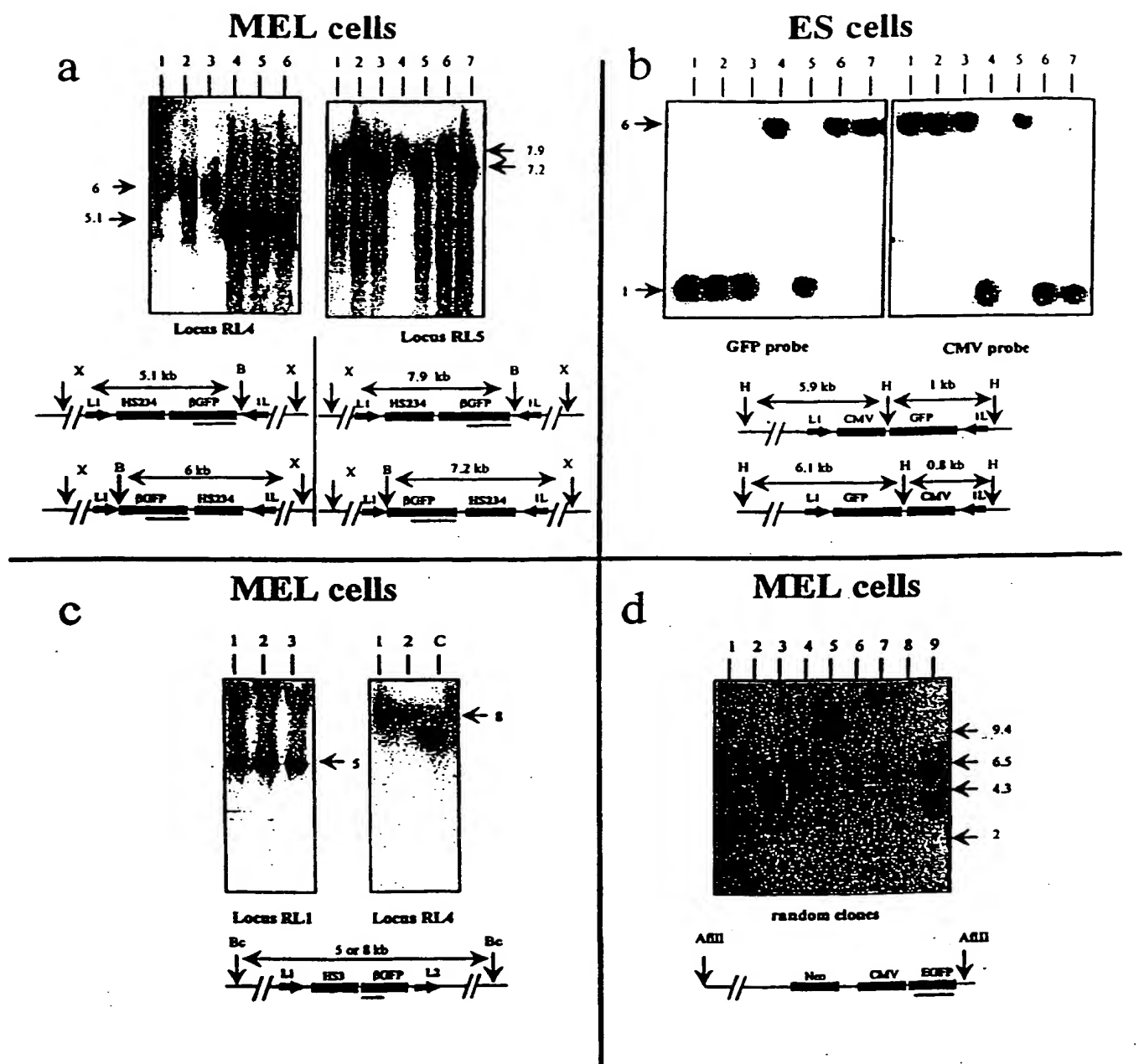


FIGURE 5

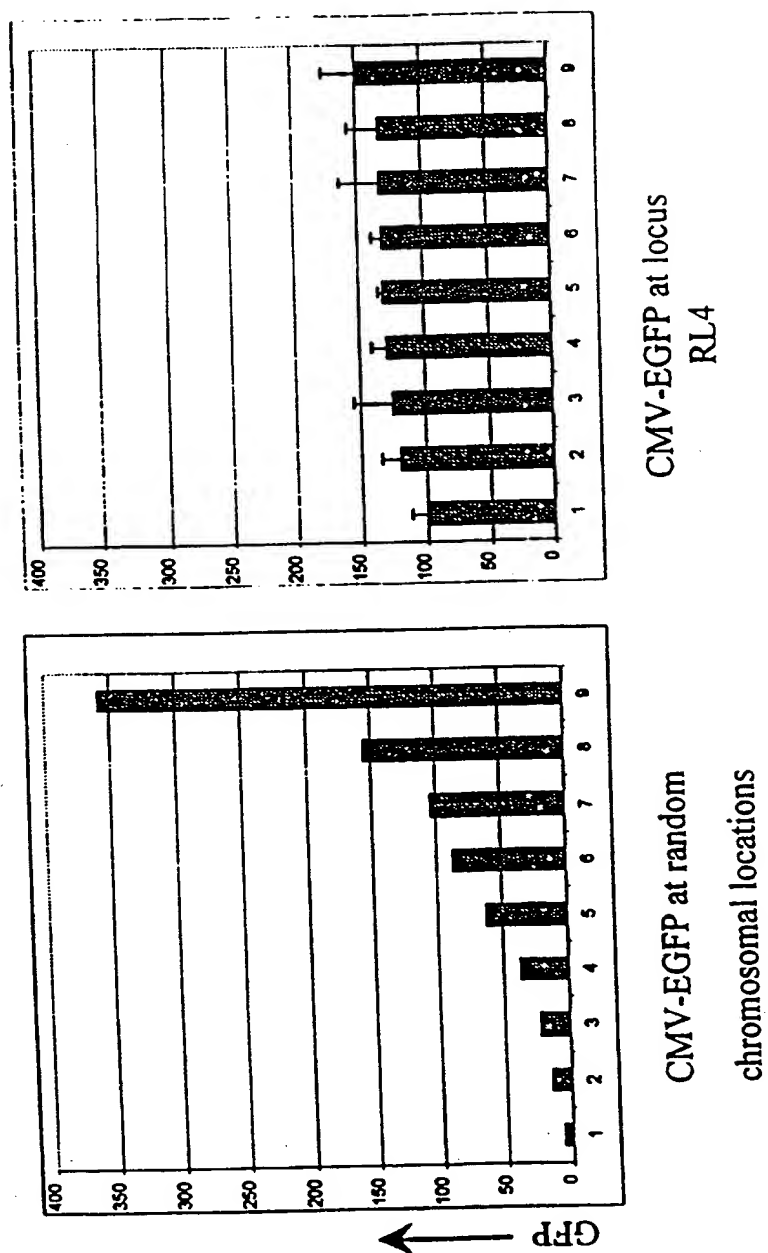


FIGURE 6

SEQUENCE LISTING

<110> MASSACHUSETTS INSTITUTE OF TECHNOLOGY and
ALBERT EINSTEIN COLLEGE OF MEDICINE OF YESHIVA UNIVERSITY

<120> METHODS AND COMPOSITIONS FOR TRANSFORMING CELLS

<130> MTZ-198CPPC

<140> 09/293,303

<141> 1999-04-16

<160> 2

<170> PatentIn Ver. 2.0

<210> 1

<211> 34

<212> DNA

<213> Artificial Sequence

<220>

<223> Description of Artificial Sequence:integration site

<400> 1

ataacttcgt ataatgtatg ctatacgaag ttat

34

<210> 2

<211> 34

<212> DNA

<213> Artificial Sequence

<220>

<223> Description of Artificial Sequence:integration site

<400> 2

ataacttcgt ataatgtata ctatacgaag ttat

34

INTERNATIONAL SEARCH REPORT

International Application No

PCT/US 00/09782

A. CLASSIFICATION OF SUBJECT MATTER
IPC 7 C12N15/90

According to International Patent Classification (IPC) or to both international classification and IPC

B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)

IPC 7 C12N

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched

Electronic data base consulted during the international search (name of data base and, where practical, search terms used)

BIOSIS

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category *	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
Y	WO 97 47758 A (EINSTEIN COLL MED ; MASSACHUSETTS INST TECHNOLOGY (US)) 18 December 1997 (1997-12-18) the whole document	1-30
Y	BOUHASSIRA E. E. ET AL.: "Transcriptional behavior of LCR enhancer elements integrated at the same chromosomal locus by recombinase-mediated cassette exchange." BLOOD, vol. 90, no. 9, 1997, pages 3332-3344, XP002143981 ISSN: 0006-4971 figure 1	1-30



Further documents are listed in the continuation of box C.



Patent family members are listed in annex.

* Special categories of cited documents:

- "A" document defining the general state of the art which is not considered to be of particular relevance
- "E" earlier document but published on or after the international filing date
- "L" document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)
- "O" document referring to an oral disclosure, use, exhibition or other means
- "P" document published prior to the international filing date but later than the priority date claimed

"T" later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention

"X" document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone

"Y" document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art.

"&" document member of the same patent family

Date of the actual completion of the international search

1 August 2000

Date of mailing of the international search report

21/08/2000

Name and mailing address of the ISA

European Patent Office, P.B. 5818 Patentlaan 2
NL - 2280 HV Rijswijk
Tel. (+31-70) 340-2040, Tx. 31 651 epo nl,
Fax: (+31-70) 340-3016

Authorized officer

Mandl, B

INTERNATIONAL SEARCH REPORT

Int. onal Application No
PCT/US 00/09782

C.(Continuation) DOCUMENTS CONSIDERED TO BE RELEVANT

Category *	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
Y	FR 2 745 008 A (ASS POUR LE DEV DE LA RECH EN) 22 August 1997 (1997-08-22) figure 1	1-30
Y	WO 97 09439 A (GENVEC INC ;KOVESDI IMRE (US); MCVEY DUNCAN L (US)) 13 March 1997 (1997-03-13) page 7, line 19 -page 11, line 8	1-30
P,X	FENG Y.-Q. ET AL.: "Site-specific chromosomal integration in mammalian cells: Highly efficient CRE recombinase-mediated cassette exchange." JOURNAL OF MOLECULAR BIOLOGY, vol. 292, no. 4, 1 October 1999 (1999-10-01), pages 779-785, XP000929622 ISSN: 0022-2836 the whole document	1-30

INTERNATIONAL SEARCH REPORT

Information on patent family members

International Application No

PCT/US 00/09782

Patent document cited in search report	Publication date	Patent family member(s)	Publication date
WO 9747758 A	18-12-1997	US 5928914 A AU 3306297 A CA 2258007 A EP 0914457 A	27-07-1999 07-01-1998 18-12-1997 12-05-1999
FR 2745008 A	22-08-1997	AU 707684 B AU 2098997 A CA 2247517 A EP 0896620 A WO 9731108 A JP 2000505298 T	15-07-1999 10-09-1997 28-08-1997 17-02-1999 28-08-1997 09-05-2000
WO 9709439 A	13-03-1997	US 5801030 A AU 717597 B AU 6912296 A BG 102344 A CN 1200149 A CZ 9800612 A EP 0850312 A HU 9802538 A NO 980838 A NZ 316806 A PL 325372 A SK 27798 A US 6063627 A	01-09-1998 30-03-2000 27-03-1997 30-04-1999 25-11-1998 15-07-1998 01-07-1998 01-02-1999 21-04-1998 28-10-1999 20-07-1998 09-09-1998 16-05-2000

CORRECTED VERSION

(19) World Intellectual Property Organization
International Bureau



(43) International Publication Date
26 October 2000 (26.10.2000)

PCT

(10) International Publication Number
WO 00/63410 A1

(51) International Patent Classification⁷: C12N 15/90

(74) Agents: REMILLARD, Jane, E. et al.; Lahive & Cockfield, LLP, 28 State Street, Boston, MA 02109 (US).

(21) International Application Number: PCT/US00/09782

(22) International Filing Date: 12 April 2000 (12.04.2000)

(25) Filing Language: English

(26) Publication Language: English

(30) Priority Data:
09/293,303 16 April 1999 (16.04.1999) US

(63) Related by continuation (CON) or continuation-in-part (CIP) to earlier application:
US 09/293,303 (CON)
Filed on 16 April 1999 (16.04.1999)

(71) Applicants (for all designated States except US): MASSACHUSETTS INSTITUTE OF TECHNOLOGY [US/US]; 77 Massachusetts Avenue, Cambridge, MA 02139 (US). ALBERT EINSTEIN COLLEGE OF MEDICINE OF YESHIVA UNIVERSITY [US/US]; 1300 Morris Park Avenue, Bronx, NY 10461 (US).

(72) Inventors; and

(75) Inventors/Applicants (for US only): BOUHASSIRA, Eric [FR/US]; 123 Farragut Avenue, Hastings on Hudson, NY 10706 (US). LEBOULCH, Philippe [FR/US]; Unit 729, 197 Eighth Street, Charlestown, MA 02129 (US).

(81) Designated States (national): AE, AG, AL, AM, AT, AU, AZ, BA, BB, BG, BR, BY, CA, CH, CN, CR, CU, CZ, DE, DK, DM, DZ, EE, ES, FI, GB, GD, GE, GH, GM, HR, HU, ID, IL, IN, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MA, MD, MG, MK, MN, MW, MX, NO, NZ, PL, PT, RO, RU, SD, SE, SG, SI, SK, SL, TJ, TM, TR, TT, TZ, UA, UG, US, UZ, VN, YU, ZA, ZW.

(84) Designated States (regional): ARIPO patent (GH, GM, KE, LS, MW, SD, SL, SZ, TZ, UG, ZW), Eurasian patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European patent (AT, BE, CH, CY, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE), OAPI patent (BF, BJ, CF, CG, CI, CM, GA, GN, GW, ML, MR, NE, SN, TD, TG).

Published:

— with international search report

(48) Date of publication of this corrected version:
31 January 2002

(15) Information about Correction:
see PCT Gazette No. 05/2002 of 31 January 2002, Section II

For two-letter codes and other abbreviations, refer to the "Guidance Notes on Codes and Abbreviations" appearing at the beginning of each regular issue of the PCT Gazette.

(54) Title: METHODS AND COMPOSITIONS FOR TRANSFORMING CELLS USING SITE-SPECIFIC RECOMBINATION

(57) Abstract: Methods and compositions for transforming cells, resulting in efficient and stable site-specific integration of transgenes, are disclosed. Transformation is achieved by providing an acceptor DNA containing two inverted lox sequences and a donor DNA containing the same two inverted lox sequences, and then contacting the acceptor and donor DNA with a recombinase (e.g., Cre or Flp) which causes recombination at the lox sequences contained in the DNAs. Prior to recombination, the acceptor DNA is preferably integrated into the genome of a cell, such as an embryonic stem cell or a fertilized egg. The acceptor DNA optionally may further contain a negatively selectable marker to allow for screening of cells which have undergone the desired site-specific recombination (e.g., DNA cassette exchange).

WO 00/63410 A1

METHODS AND COMPOSITIONS FOR TRANSFORMING CELLS USING SITE-SPECIFIC RECOMBINATION

Government Funding

5 This invention was made with government support under grants HL38655, HL 554350, DK54071-020 and HL07556, all of which were awarded by the NIH. The United States government has certain rights in the invention.

Background of the Invention

10 The success of gene therapy techniques depends largely on the ability to achieve a combination of stable chromosomal integration and high-level, regulated expression of transferred genes. Regulated gene expression is most easily achieved by means of large DNA fragments containing extensive cis-acting regulatory regions. For example, gene therapy for β -globin disorders may require high-level, position-independent expression
15 of extended gene and LCR sequences.

 Many current techniques allow efficient transient transfection of cells *in vitro* and *in vivo* with large DNA fragments. However, subsequent chromosomal integration is very inefficient. To overcome low levels of integration, retroviral vectors which integrate very efficiently in permissive cells can be used. However, such vectors are
20 greatly limited by constraints of size and sequence composition.

 There are also many other techniques available for stable integration of transgenes in mammalian cells (Kriegler, M. (1990) *Gene Transfer and Expression. A Laboratory Manual*, Stockton Press, New York); and (Wolf, J.A. (1994) *Gene Therapeutics: Methods and Application of Direct Gene Transfer*, Birkhauser, Boston)..
25 However, these methods result in integration at random chromosomal locations of an uncontrolled number of transgene copies that express at levels that generally cannot be predicted or reproduced with precision because of position-effects. The inability to control the site of integration, the number of integrated copies and the level of expression of transgenes has impeded progress in studies of both gene expression and
30 the physiological effects of transgenes.

Systems which can perform site-specific chromosomal integration efficiently therefore have wide utility. The first site-specific chromosomal integrations in mammalian cells were based on integration of a single Lox or FRT site on a chromosome followed by trapping of rare integration events (O'Gorman et al. (1991) *Science* 251:1351-1355; and Sauer, B. (1994) *Current Opinion in Biotechnology* 5:521-527). These pioneering methods had three limitations: 1) they were quite inefficient, 2) the entire plasmid was integrated, and 3) a positive selectable marker was left in the chromosome after the integration. The low efficiency of these methods is due to the reversibility of the recombination reaction: after integration the transgene is re-excised if the two identical Lox or FRT sites that flank the transgene recombine with each other. Since the excision reaction is intra-molecular while the insertion reaction is inter-molecular, excisions are favored.

US 4,959,317 discloses the use of Cre-Lox site-specific recombination to achieve gene transfer in eukaryotic cells (Sauer et al. (1993) *Methods in Enzymology* 225: at 898). The target site of the CRE recombinase is a 34 bp sequence that consists of two inverted 13 bp CRE-binding sites separated by an eight base spacer within which the recombination occurs (Hoess, et al. (1984) *Proc. Nat. Acad. Sci. (USA)* 81:1026-1029).

Additional site-specific DNA recombination systems which provide more efficient and stable integration of transgene sequences into genomic DNA, preferably without the use of a positively selectable marker, would be greatly beneficial.

Summary of the Invention

The present invention provides methods and compositions for achieving efficient and stable site-specific DNA recombination using a recombinase/lox system, such as the Cre/lox system or the Flp/rt system. In one embodiment, the method comprises contacting a recombinase (e.g., Cre or Flp) with (a) an acceptor vector comprising two incompatible lox sequences, L1 and L2, and (b) a donor vector comprising a selected DNA flanked by the L1 and L2 sequences, or sequences which are compatible with the L1 and L2 sequences, thereby causing transfer of the selected DNA from the donor vector into the acceptor vector by recombination at the compatible lox sequences. In a

preferred embodiment, the acceptor vector is a retroviral vector or an adeno-associated vector.

In another embodiment, the invention provides a method of transforming a cell with a selected DNA comprising, in any order, the steps of introducing into the cell a
5 acceptor vector which integrates into the genome of the cell, the acceptor vector comprising two incompatible lox sequences, L1 and L2, (b) introducing into the cell a donor vector comprising the selected DNA flanked by the L1 and L2 sequences, or sequences which are compatible with the L1 and L2 sequences, and (c) contacting L1 and L2 with a recombinase, such as Cre or Flp, thereby causing transfer of the selected
10 DNA from the donor vector into the acceptor vector. The recombinase can be introduced into the cell in the form of a protein or a gene encoding the protein.

In another embodiment, the invention provides a vector selected from the group consisting of retroviral vectors and adeno-associated vectors comprising two incompatible lox sequences, L1 and L2.

15 In another embodiment, the invention provides a method of achieving site-specific recombination by providing a donor DNA comprising two inverted lox sequences, and an acceptor DNA comprising the same two inverted lox sequences contained in the donor DNA, and then contacting the donor and acceptor DNA with a recombinase (e.g., Cre or Flp). Preferably, the acceptor DNA is integrated into the
20 genome of a host cell prior to contact with the recombinase (e.g., by homologous recombination), so that recombination results in site-specific genomic integration of a desired transgene or other polynucleotide. In another preferred embodiment, the donor DNA is present in excess of the acceptor DNA. Suitable lox sequences comprise the nucleotide sequences of SEQ ID NO:1, SEQ ID NO:2 and inversions thereof. The
25 donor and acceptor DNAs can further contain selectable markers which are generally positioned between the two inverted lox sequences to help select for cells which have undergone the desired recombination.

The methods and compositions of the invention can be used in methods of *in vivo* and *In vitro* gene transfer (e.g., gene therapy) to cause efficient and stable site-
30 specific (targeted) integration of transgene sequences. By controlling the site or position of integration of a transgene into the genome of a cell, expression levels can be

predictably determined. For example, in cells, the invention can be used to produce desired proteins (e.g., drugs) by insertion of transgenes at pre-selected chromosomal locations where expression of the transgene will be high. Similarly, the invention can be used to develop "designer proteins" by insertion of multiple versions of a gene or DNA (e.g., related variants) at the same locus to test the various versions of proteins produced in a context in which the proteins are all produced in the same amount. The invention can also be used to study and to identify genetic elements that control position effects.

In addition, the invention can be used *in vivo* to create transgenic mammals and/or plants. For example, animal models of human disease can be generated, particularly if multiple genes have to be expressed at well regulated levels. Animals and/or plants can be created which contain target lox sequences (e.g., inverted or incompatible) at chromosomal locations that are not subject to position effects or to desired position effects for directing expression of a gene of interest. This allows for the generation of animal models and/or plants with, for example, higher resistance to disease or improved physical/functional characteristics.

Brief Description of the Figures

Figure 1 is a schematic illustration of Cre/lox mediated gene transfer. Lox A and Lox B are mutually incompatible lox sites which are unable to recombine with each other in the presence of Cre recombinase.

Figure 2 is an illustration of acceptor and donor vectors containing selectable marker genes and incompatible lox sequences. Following pre-integration of the acceptor vector into a genome, the donor vector undergoes site-specific Cre-mediated recombination with the acceptor vector at the lox sequences, resulting in stable DNA exchange between the vectors

Figure 3 is an illustration of a Cre expression vector and a control Cre expression vector.

Figure 4 is a schematic illustration of recombinase-mediated DNA cassette exchange using either incompatible exchange sequences (e.g., lox L1 and L2) or inverted exchange sequences (e.g., lox L1 and 1L). **Panel A** shows how site-specific chromosomal integration can be performed by exchanging a pre-integrated chromosomal cassette (cas1) with a plasmidic cassette (cas 2), using incompatible lox sequences. Recombination through the L1 site leads to plasmid integration; recombination through L2 leads to resolution of the array thus created. The end result of the recombination reaction is therefore an exchange of cassette 1 by cassette 2. This exchange is stable because the cassette is flanked by Lox sites that cannot recombine with each other.

10 Symmetrical integration through L2 and resolution through L1 also lead to the cassette exchange. After integration the resolution step is in competition with an excision step that is the reverse of the integration step. **Panel B** shows how selection for the loss of the HYTK cassette can be used to identify clones which have undergone site-specific chromosomal recombination. **Panel C** shows site-specific chromosomal recombination using inverted Lox sites: The L2 Lox site is replaced by an inverted L1 site (1L). Excision products are eliminated because recombination between Lox sites located in cis leads to inversion of the HYTK gene rather than its excision (upper recombination). Gancyclovir selection therefore lead predominantly to the identification of exchange products (lower recombination). Exchange products in both possible orientations are

20 obtained. Only one of several exchange pathways is depicted. **Panel D** shows FACS-based site-specific chromosomal recombination. A plasmid containing the CMV-EGFP reporter gene inserted outside of the exchange cassette is co-transfected (with a CRE expression plasmid) and cells that express the highest amount of GFP are sorted. Among those cells, exchange frequency is relatively high. No selectable marker on the

25 chromosome is required before or after the exchange.

Figure 5 shows Southern blots confirming site-specific chromosomal recombination in MEL and ES cells. **Panel A** is an autoradiogram illustrating the exchange of the HYTK gene with the HS234 β GFP cassette at two loci in MEL cells.

30 Lanes 1 to 6 in the left panel, and lanes 1 to 7 in right panel, contain genomic DNA from Gan^R resistant clones digested with EcoRV and Bgl II and probed with a fragment

encompassing the EGFP coding sequence. Two bands (6.1 and 5.1 at RL4 or 7.9 and 7.2 kb at RL5) were observed per locus demonstrating insertion in both possible orientations. B= BglII; X= EcoRV or BglII.

Panel B is an autoradiogram illustrating the exchange of the HYTK gene with the HS234 β GFP cassette at two loci in ES cells. Lanes 1 to 7 in the left panel contain genomic DNA from Gan^R resistant clones digested with HindIII and probed with a fragment encompassing the EGFP coding sequence. The right panel is a blot of the left panel stripped and re-probed with a fragment encompassing the CMV promoter. As expected, clones in both orientations were obtained. **Panel C** shows recombinase-mediated DNA cassette exchange using no selectable marker. Left panel, lanes 1 to 3, and right panel, lanes 1 and 2, contain genomic DNA digested with BclI and probed with a fragment encompassing the human β -globin promoter; lane C contain control DNA. As expected, all exchanges were in the same orientation. Bc= BclI. **Panel D** shows random (uncontrolled) integration of Cassette CMV-EGFP: Lanes 1 to 9 contain genomic DNA digested with AflII probed with the EGFP coding sequence. Bands of all sizes and intensity characteristic of uncontrolled integration are shown.

Figure 6 is a chart showing expression levels of GFP at pre-selected locus and at a random locus: GFP expression of nine independent clones with the CMV-EGFP cassette integrated at 9 uncontrolled random loci (left panel) or at the same pre-selected locus (right panel) was analyzed by FACS. The y-axis represents the mean linearized GFP fluorescence. The results show that site-specific integration considerably diminishes the variability of expression typically observed in stable transfection studies.

25 Detailed Description of the Invention

The present invention provides methods and compositions for causing efficient site-specific DNA recombination, for example, in cells or fertilized eggs. The advantages over current cell transformation techniques provided by the invention include highly efficient and stable integration of large DNA sequences, such as transgene sequences, into chromosomal DNA, without the need for a selectable marker in the integrated DNA. In one embodiment, the method of the invention comprises

contacting a recombinase, such as Cre, with (a) an acceptor DNA (e.g., vector) comprising two incompatible lox sequences, L1 and L2, and (b) a donor DNA (e.g., vector) comprising a selected DNA flanked by the L1 and L2 sequences, or lox sequences which are compatible with the L1 and L2 sequences, thereby causing transfer
5 of the selected DNA from the donor vector into the acceptor vector by recombination at the compatible lox sequences.

In another embodiment, the method of the invention comprises contacting a recombinase, such as Cre or Flp, with (a) a donor DNA comprising two inverted lox sequences, and (b) an acceptor DNA comprising the same two inverted lox sequences
10 contained in the donor DNA. Preferably, the acceptor DNA is integrated into the genome of a host cell (e.g., by homologous recombination), prior to contact with the recombinase so that recombination results in site-specific genomic integration of a desired transgene or other polynucleotide.

The term "site-specific recombination," refers to DNA transfer from a donor
15 DNA or vector to an acceptor DNA or vector.

The term "lox sequence" refers to a nucleotide sequence which undergoes recombination (e.g., DNA cross-over and exchange) when catalyzed by a recombinase, such as Cre, Flp or another member of the Int family of recombinases (Argos et al. (1986) *EMBO J.* 5: 433). Suitable lox sequences include, for example, the lox
20 sequences recognized by Cre recombinase, and the *frt* sequences recognized by Flp recombinase.

The term "recombinase" refers to any recombinase capable of catalyzing a site-specific recombination at a lox site. Suitable recombinases include, for example, Cre recombinase (Sauer et al. (1993) *Methods in Enzymology* 225: 898) and Flp recombinase
25 (Buchholz et al. (1996) *Nucl. Acids Res.* 24:4256-4262; Buchholz et al. (1998) *Nat. Biotechnol.* 16:657-662).

The term "incompatible lox sequences" refers to two or more lox sequences (referred to herein as L1 and L2) which differ from one another and, therefore, can not undergo recombination with one another. For example, lox sequences can be rendered
30 incompatible if their nucleotide sequences differ by even one nucleotide, particularly in

their spacer regions. In contrast, the term "compatible lox sequences" refers to two or more lox sequences which can recombine when catalyzed to do so by a recombinase.

The term "inverted lox sequences" refers to two or more substantially identical lox sequences (referred to herein as L1 and 1L) which are positioned in the reverse orientation with respect to one another so that recombination between the L1 and 1L sequences within the same (e.g., acceptor) DNA (e.g., intrachromosomal recombination) leads to an inversion of the intervening sequence, as opposed to an excision (see Figure 4). "Inverted lox sequences" of the invention recombine with the same lox sequences (e.g., contained within a separate donor DNA) at the same frequency (i.e., specific activity) because they are made up of identical or substantially identical nucleotide sequences. Overall, these features provide a site-specific recombination system which results in (1) substantially equal amounts of intermolecular recombination between identical L1 or 1L sequences (e.g., contained on separate acceptor and donor DNAs), and (2) substantially no excision from intramolecular recombination between inverted L1 and 1L sequences (e.g., contained within the same acceptor or donor DNA). As shown schematically in Figure 4, this allows for highly (up to 100%) efficient and stable recombinase-mediated DNA integration.

The terms "acceptor DNA" and "acceptor vector," are used interchangeably herein and refer to any DNA or vector which, preferably, is capable of integrating into the genome of a cell. For example, the acceptor DNA or vector can be of viral origin, such as a retroviral vector or adeno-associated vector. Generally, the acceptor DNA or vector contains an exchange cassette (i.e., DNA which is replaced by DNA from the donor vector) and can also, optionally, contain a selectable (e.g., negative) marker gene.

The terms "donor DNA" and "donor vector" are used interchangeably herein and refer to any DNA or vector (e.g., circular plasmid DNA) containing DNA which is transferred to the acceptor DNA or vector via a recombinase-mediated exchange reaction. Generally, the donor DNA or vector comprises plasmid DNA and, optionally, also can contain a selectable (e.g., positive) marker gene.

The method of the present invention utilizes a recombinase-mediated exchange reaction which takes place between identical or compatible (i.e., able to recombine with one another) lox sequences. The efficient exchange of DNA between identical or

compatible lox sequences enables transfer of DNA from the donor to the acceptor vector, which each contain identical or compatible lox sites (see Figure 4). However, once transferred from donor to acceptor vector (i.e., intermolecular transfer), the transferred DNA is stabilized or "locked" into place. In one embodiment, this is achieved by using incompatible lox sequences (e.g., L1 and L2) within the acceptor vector which prevent intramolecular exchange and excision of the transferred DNA. In another embodiment, this is achieved by using identical inverted lox sequences (e.g., L1 and 1L) within the acceptor vector which recombine with each other in reverse so that the transferred DNA is inverted within the vector without excision of the transferred DNA. Therefore, the transferred DNA is integrated in a highly efficient and stable manner.

In addition to effective and stable DNA exchange reactions, the methods of the present invention take advantage of highly efficient integration vectors, such as retroviral vectors, adeno-associated vectors, or vectors encoding retroviral integrases, for use in integrating acceptor DNA into the genome of a cell. The studies described herein demonstrate that such vectors are compatible for use with site-specific DNA transfer systems, such as recombinase/lox systems.

The invention also employs an excess of donor DNA relative to acceptor DNA so that intermolecular exchange leading to insertion of donor DNA into (e.g., genomically integrated) acceptor DNA occurs at a higher frequency than intramolecular exchange (e.g., within acceptor DNA) leading to excision of donor DNA.

Overall, the site-specific recombination system of the invention provides a means for highly efficient and stable DNA transfer which can be used, for example, in methods of gene therapy, gene expression control, and transgenic applications. For example, as described herein, the methods and compositions of the present invention can be used to achieve a 100 to 10,000 fold increase in transgene integration and expression compared to random integration which occurs in the absence of site-specific recombination. The invention further provides a way to achieve site-specific recombination without the incorporation of selectable markers into the genome following exchange.

Accordingly, in another embodiment, the invention provides a method of transforming a selected cell or tissue, such as a mammalian cell (e.g., embryonic stem cell) or fertilized egg, with a desired DNA. The method comprises the steps of (a) introducing into the cell or tissue an acceptor vector comprising two incompatible lox sequences, L1 and L2, (b) introducing into the cell a donor vector comprising the selected DNA flanked by the L1 and L2 sequences, or lox sequences which are compatible with the L1 and L2 sequences, and (c) contacting L1 and L2 with a recombinase, such as Cre, to cause transfer of the selected DNA from the donor vector into the acceptor vector (by way of an exchange reaction between the compatible lox sequences). While not essential, the acceptor vector is preferably introduced into the cell prior to introduction of the donor vector, so that the acceptor vector has integrated into the host genome prior to DNA exchange with the donor vector.

Alternatively, selected cells and/or tissues can be transformed by (a) introducing into the cell or tissue an acceptor DNA comprising two inverted lox sequences; (b) introducing into the cell or tissue a donor DNA comprising the same two inverted lox sequences contained in the acceptor DNA; and (c) contacting the donor and acceptor DNA with a recombinase. The recombinase can be introduced into the cell or tissue in the form of an exogenous protein or in the form of a gene encoding the recombinase which is expressed in the cell or tissue following transfection. Suitable lox sequences again include those having the nucleotide sequences shown in SEQ ID NOS: 1 and 2, as well as inversions thereof.

The acceptor DNA can be any DNA (e.g., plasmid vector) capable of being taken up by cells and integrating into genomic DNA. Suitable acceptor DNAs include viral vectors which transfect cells directly, such as recombinant retroviruses, adenovirus, adeno-associated virus, and herpes simplex virus-1. A prerequisite for the use of retroviruses is to ensure the safety of their use, particularly with regard to the possibility of the spread of wild-type virus in the cell population. The development of specialized cell lines (termed "packaging cells") which produce only replication-defective retroviruses has increased the utility of retroviruses for gene therapy, and defective retroviruses are well characterized for use in gene transfer for gene therapy purposes (for a review see Miller, A.D. (1990) *Blood* 76:271). Thus, recombinant retrovirus can be

constructed in which part of the retroviral coding sequence (*gag*, *pol*, *env*) has been replaced by nucleic acid encoding a mutated subunits of the mALDH of the invention rendering the retrovirus replication defective. The replication defective retrovirus is then packaged into virions which can be used to infect a target cell through the use of a helper virus by standard techniques. Protocols for producing recombinant retroviruses and for infecting cells *in vitro* or *in vivo* with such viruses can be found in Current Protocols in Molecular Biology, Ausubel, F.M. et al. (eds.) Greene Publishing Associates, (1989). Sections 9.10-9.14 and other standard laboratory manuals. Examples of suitable retroviruses include pLJ, pZIP, pWE and pEM which are well known to those skilled in the art. Examples of suitable packaging virus lines for preparing both ecotropic and amphotropic retroviral systems include ψ Crip, ψ Cre, ψ 2 and ψ Am. Retroviruses have been used to introduce a variety of genes into many different cell types, including neural cells, epithelial cells, endothelial cells, lymphocytes, myoblasts, hepatocytes, bone marrow cells, *in vitro* and/or *in vivo* (see for example Eglitis, et al. (1985) *Science* 230:1395-1398; Danos and Mulligan (1988) *Proc. Natl. Acad. Sci. USA* 85:6460-6464).

Another suitable acceptor vector is an adenovirus-derived vector. The genome of an adenovirus can be manipulated such that it encodes and expresses a gene product of interest but is inactivated in terms of its ability to replicate in a normal lytic viral life cycle. See for example Berkner et al. (1988) *BioTechniques* 6:616; Rosenfeld et al. (1991) *Science* 252:431-434; and Rosenfeld et al. (1992) *Cell* 68:143-155. Suitable adenoviral vectors derived from the adenovirus strain Ad type 5 dl324 or other strains of adenovirus (e.g., Ad2, Ad3, Ad7 etc.) are well known to those skilled in the art.

Yet another viral vector system useful as the acceptor vector is the adeno-associated virus (AAV). Adeno-associated virus is a naturally occurring defective virus that requires another virus, such as an adenovirus or a herpes virus, as a helper virus for efficient replication and a productive life cycle. (For a review see Muzyczka et al. *Curr. Topics in Micro. and Immunol.* (1992) 158:97-129). It is also one of the few viruses that may integrate its DNA into non-dividing cells, and exhibits a high frequency of stable integration (see for example Flotte et al. (1992) *Am. J. Respir. Cell. Mol. Biol.* 7:349-356; Samulski et al. (1989) *J. Virol.* 63:3822-3828; and McLaughlin et al. (1989) *J.*

Virol. 62:1963-1973). Vectors containing as few as 300 base pairs of AAV can be packaged and can integrate.

Other viral vector systems that may be used as the acceptor vector in the methods of the present invention include herpes virus, vaccinia virus, and several RNA viruses.

5 Alternatively, the acceptor DNA can be targeted for delivery to selected cells (see e.g., U.S. 5,166,320) via cell-specific targeting ligands. Once taken up by cells, the acceptor DNA can be integrated into genomic DNA either randomly or selectively, for example, by homologous recombination as is well known in the art (see e.g., 5,614,396).

The acceptor DNA also can be introduced, for example, by microinjection and/or
10 homologous recombination, into a host cell, embryonic stem cell, tissue, or fertilized egg. The donor DNA can be any DNA (e.g., circular DNA) capable of being taken up by cells, either *in vivo* or *in vitro*, and capable of carrying the desired transfer (i.e., donor) DNA sequence which, for example, encodes a therapeutic gene. Suitable donor vectors include cosmids or DNA plasmids, such as recombinant bacterial or eukaryotic
15 plasmids. The donor vector can be introduced into the host cell either *in vivo* or *in vitro* using a variety of known methods. For *in vitro* delivery, suitable methods include direct injection of the plasmid (e.g., microinjection), CaPO_4 precipitation, electroporation, cationic lipofection, or use of artificial viral envelopes. For *in vivo* delivery, suitable methods include intravenous, intraperitoneal and intramuscular injection of the vector.
20 The vector can also be targeted for delivery to selected cells (see e.g., US 5,166,320).

In a preferred embodiment of the invention, the donor DNA is introduced into the host cell in excess of the acceptor DNA to increase the frequency of DNA integration relative to excision. Integration is an intermolecular event, whereas excision is an intramolecular event. Therefore, excision naturally occurs more often than
25 integration. To tilt the balance toward integration, the amount of donor DNA can be increased to increase the number of intermolecular exchange reactions which occur in the host cell.

In another embodiment, the acceptor and/or donor DNA also includes a selectable marker which enables the identification of clones which have undergone site-
30 specific recombination resulting in an exchange of acceptor DNA for donor DNA. For example, the acceptor DNA can include a negatively selectable marker gene to identify

cells which have lost the ability to produce a certain protein (e.g., which have lost resistance to a drug or antibiotic). Similarly, the donor DNA can include a positively selectable marker to identify cells which have incorporated the donor DNA into their genome via recombinase-mediated exchange. The marker gene can be promoterless so that it will only be expressed when integrated into the acceptor vector containing a promoter to drive its expression.

The donor and acceptor vectors each contain at least two incompatible lox sequences ("L1 and L2") so that intramolecular recombination can not occur. At the same time, the locks sequences of the donor and acceptor vectors must be able to recombine intermolecularly (e.g., L1 with a compatible L1, and L2 with a compatible L2) with one another to allow DNA exchange between the donor and acceptor vectors. In order to ensure intermolecular exchange between compatible lox sequences, the lox sequences are generally oriented in the same direction.

Incompatibility between locks sequences can be achieved, for example, by way of mutating or modifying (e.g., by nucleotide addition, deletion or substitution) one of two identical lox sequences, preferably in their spacer sequences, so that the sequences differ. Testing to determine which mutations confer incompatibility can be performed using standard mutation assays which test for the ability of the mutated and non-mutated lox sequences to recombine.

In a preferred embodiment, one of the two incompatible lox sequences is the Lox P1 sequence of the Cre/lox system of bacteriophage P1 (Hoess et al. (1990) "Nucleic Acids and Molecular Biology," Vol 4, p. 99) having the sequence shown in SEQ ID NO: 1. The Lox P1 sequence is a 34 base pair sequence which can be isolated from bacteriophage P1 by methods known in the art (see e.g., Hoess et al. (1982) *PNAS* 79:3398). The Lox P1 sequence consists of two 13 base pair inverted repeats separated by an eight base spacer sequence. Lox P1 sites can also be isolated from plasmids available from the ATCC (e.g., ATCC 53254 and 20773). Other suitable lox sequences include the Lox B, Lox L, and Lox R sequences isolatable from *E. coli* (Hoess et al. (1982), *supra.*). Lox sequences can also be chemically synthesized using known techniques, such as those described in the Examples below.

Accordingly, the other incompatible lox sequence can be a mutated form of the LoxP1 sequence, for example, having a point mutation in the eight nucleotide spacer sequence. In one embodiment, the point mutation is substitution of A for G at position 7 of the eight base spacer sequence of the wild type Lox P1 sequence, referred to herein as the Lox511 sequence (SEQ ID NO: 2). Accordingly, in one embodiment, the two incompatible lox sequences of the invention have the following sequences:

		<u>SPACER</u>
	Lox P1	ATAACTTCGTATA ATGTAT <u>G</u> C TATACGAAGTTAT
10	Lox 511	ATAACTTCGTATA ATGTATA <u>A</u> C TATACGAAGTTAT

Alternatively, the donor and acceptor DNAs (e.g., vectors) contain two identical but inverted lox sequences so that intramolecular recombination between the inverted lox sequences results in an inversion of the intervening DNA, and intermolecular recombination between identical lox sequences results in exchange of donor and acceptor DNA (see Figure 4). Moreover, because the lox sequences involved are all substantially identical, the frequency of recombination among them is the same. These features provide highly (up to 100%) efficient and stable recombinase-mediated DNA integration. In fact, in one embodiment, cells which have taken up the largest amount of donor DNA, as measured by e.g., a transient, non-integrated marker gene (e.g., green fluorescence protein), can be selected for and, among these cells, DNA exchange will be high enough that no selectable marker need be used. In these embodiments, as in those described in the preceding paragraphs, suitable inverted lox sequences include but are not limited to the LoxP1 (SEQ ID NO:1), Lox511(SEQ ID NO:2) and inversions thereof.

Intermolecular recombination between compatible or inverted lox sequences in the donor and acceptor vectors is catalyzed by a recombinase, such as Cre or another member of the Int family of recombinases (Argos et al. (1986) *EMBO J.* 5: 433) which have been shown to perform efficient recombination at lox sequences in both bacteria and in eukaryotic cells (Sauer et al. (1993) *Methods in Enzymology* 225: 890-900). The recombinase can be introduced into the cell along with the donor and acceptor vectors in

the form of a protein or as an expressible gene encoding the protein (e.g., the Cre gene described by Sauer, B. et al. (1988) *Proc. Natl. Acad. Sci. USA* 85:5166-5170). The recombinase or recombinase gene can be introduced or transfected into the host cell before, simultaneously, or following introduction of the donor and acceptor vectors.

5 In one embodiment, the recombinase gene (e.g., Cre) is contained in an expression vector which is co-transfected with the donor vector following introduction and integration of the acceptor vector into the host cell. In another embodiment, the recombinase gene is contained within either the acceptor vector or the donor vector. As with the donor vector, the recombinase gene can be introduced into the host cell either *in*
10 *vivo* or *in vitro* using known techniques, such as CaPO₄ precipitation, electroporation, cationic lipofection, use of artificial viral envelopes, microinjection (e.g., into fertilized eggs), direct injection (e.g., intravenous, intraperitoneal or intramuscular). The vector can also be targeted for delivery to selected cells (see e.g., US 5,166,320).

The DNA which is transferred from the donor to the acceptor vector by way of
15 the site-specific recombination method of the invention can be any DNA desired for stable integration into a host cell genome. For example, the gene can be any transgene useful, for example, in gene therapy or for diagnostic purposes. The gene can encode a desired therapeutic protein, such as α , β or δ globin, blood coagulation factors (e.g., Factors VIII and IX) gene, cell surface receptors and other desirable proteins, for
20 example, to correct inherited deficiencies of these proteins in an individual.

Accordingly, in one embodiment, the invention can be used *in vivo* to create transgenic mammals and/or plants. For example, animals of human disease can be generated, particularly if multiple genes have to be expressed at well regulated levels. Transgenic animals and/or plants can be created using known techniques, such as
25 microinjection or other methods of introducing DNA into fertilized eggs or cells (e.g., embryonic stem cells), which contain target lox sequences (e.g., inverted or incompatible) at chromosomal locations that are not subject to position effects or to desired position effects for directing expression of a gene of interest. This allows for the generation of animals and/or plants with, for example, higher resistance to disease or
30 improved physical/functional characteristics.

In general, there are two broad types of applications for the invention depending on whether the Lox sites are integrated in the genome at random sites or at known chromosomal locations (e.g., via homologous recombination). Lox sites at random locations can be used to generate reference loci, such as loci with characteristic position-effects. For instance, the invention can be used to site-specifically integrate particular genes into reference loci favorable for strong expression in a particular tissue, or loci that confer a pan-cellular or a variegated pattern of expression. Moreover, site-specifically integrated donor DNA at known genomic locations can facilitate the systematic characterization of endogenous cis-regulatory elements and the genetic dissection of protein function via the generation of mutated versions of endogenous gene products.

Accordingly, in other embodiments, the invention provides a means for (1) the discovery and analysis of cis-acting DNA regulatory elements controlling expression and position-effects 2) the production of polypeptides of biological value by insertion of transgenes at reference loci at which expression is high or optimally inducible, (3) the development of designer proteins by insertion of multiple versions of a gene at the same locus in order to test variants in a context in which they are produced in the same amount and have the same tissue distribution, (4) the creation of animal models of human diseases (particularly those in which multiple genes must be expressed at well regulated levels), and (5) the creation of improved agricultural organisms.

Classical gene transfer technology often involves cumbersome screening procedures to identify clones or animals with appropriate levels and developmental patterns of expression. These first generation technologies are sufficient to perform simple genetic manipulations, however, they are inadequate for the implementation of complex genetic strategies to modify plants and animals by introduction of large numbers of carefully regulated transgenes to achieve a desired phenotype. For such complex genetic engineering purposes the development of rules for the creation of artificial genetic loci (AGL) is required.

The compositions and methods of the present invention provide a means for creating an AGL by allowing for the production of an array of genes and regulatory elements integrated at reference genomic sites whose chromatin structure and influence on gene

expression can be controlled at will by insertion of known cis-acting elements. Thus, transgenes making up an AGL can be expressed in a tissue and in a developmental stage-specific manner at levels that are totally predictable. The present invention facilitate the development of AGLs by both permitting the careful characterization of the site with different cis-regulatory elements and mediating the efficient and convenient exchange of multiple transgenes into the AGL. Accordingly, the methods and compositions of the invention can be used for a variety of therapeutic and diagnostic applications which require stable and efficient integration of transgene sequences into genomic DNA of cells. The methods and compositions can be used to transform a wide variety of eukaryotic cells (e.g., mammalian) cells and provide the advantage of high efficiency DNA transfer.

EQUIVALENTS

Those skilled in the art will recognize, or be able to ascertain using no more than routine experimentation, many equivalents to the specific embodiments of the invention described herein. Such equivalents are intended to be encompassed by the following claims. The entire contents of all references, published patent applications and issued patents cited throughout this application are hereby incorporated by reference.

EXAMPLES

EXAMPLE 1 - CRE/LOX MEDIATED GENE INTEGRATION USING INCOMPATIBLE MUTATED LOX SEQUENCES

DNA Construction and Cell Culture

DNA vectors were made using standard techniques (Sambrook, J. et al. (1989) *Molecular Cloning: A Laboratory Manual - 2nd ed.* Cold Spring Harbor Laboratory Press, Cold Spring Harbor, New York, USA). Oligonucleotides were synthesized by Research Genetics, Inc. Accuracy of DNA construction was verified by sequencing. LXSN retroviral vector (Miller, A.D. et al. (1989) *Biotechniques* 7:980-990) was provided by D. Miller (Fred Hutchinson Cancer Research Center, Seattle), Hygromycin B (Lupton, S.D. et al. (1991) *Mol. Cell. Biol.* 11:3374-3378) phosphotransferase gene by

D. Housman (MIT, Cambridge). Herpes Simplex virus thymidine kinase (HSV-TK) gene (Lupton, *supra*) by M. R. Capecchi (Salt Lake City, Utah), U19 SV40T mutant gene (Renfranz, P.J. et al. (1991) *Cell* 66:713-729) by R. D. McKay (MIT, Cambridge) and G. Almazan (McGill University, Montreal), Cre recombinase gene (Sauer, B. et al. (1988) *Proc. Natl. Acad. Sci. USA* 85:5166-5170) by D. W. Ow (UC Berkeley, Albany), CD24 (Pawlink, R. et al. (1994) *Blood* 84: 2868-2877). MSCV (murine stem cell virus) retroviral vector (Hawley, P.G. et al. (1994) *Gene Therapy* 1:136-138), pBabe retroviral vector (Morgenstern, J. P. et al. (1990) *Nucl. Acids Res.* 18:3587-3596) by R. Weinberg (MIT, Cambridge), pcDNA1 by Invitrogen Corp. and pOPRSVICAT by Stratagene, Inc. NIH3T3 cells were obtained from the ATCC, BOSC23 cells (Pear, W. S., et al. ((1993) *Proc. Natl. Acad. Sci., USA* 90:8392-8396) W. Pear and D. Baltimore (Rockefeller University, New York).

NIH3T3 cells were grown at 37°C with 5% CO₂/95% air in DMEM supplemented with 10% heat inactivated calf serum (CS), 4.5 mg/ml glucose, 2 mM glutamine, 100 IU/ml penicillin and 100 µg/ml streptomycin. For BOSC23 cells, CS was replaced by 10% heat inactivated fetal calf serum (FCS).

Cell Infection, Transfection and Selection

The packaging cell line, BOSC23, was grown as described (Pear, *supra*, Danos, O. et al. (1988) *Proc. Natl. Acad. Sci. USA* 85:6460-6464). Plasmid DNA's were prepared by the Qiagen procedure (Qiagen, Inc.) and transfected in BOSC23 cells using a calcium phosphate procedure (5prime:3prime, Inc.). Viral supernatants from producers were harvested and filtered as described (Pear, *supra*, Danos, *supra*). All infections were carried out in the presence of 8 µg/ml Polybrene (Sigma). Viral supernatants from BOSC23 were used to generate stable viral producers. Virus titers were estimated by infection and selection of NIH3T3 cells using standard calculations previously described (Pear, *supra*, Danos, *supra*). Detection of helper viruses was performed by a β-galactosidase mobilization assay as described (Pear, *supra*, Danos, *supra*). Selection was applied two days after infections.

Standard concentrations (1 X) of selection agents were 320 µg/ml for Hygromycin B (Calbiochem). Packaging NIH3T3 cells were selected with 1X, MDHF with 1/2 X and BSMC with 2 X concentrations.

5 Site Specific DNA Integration Using Incompatible Mutated Lox Sequences

To study the efficiency of gene integration using the Cre/lox mediated gene transfer system described herein, the following protocol was performed.

10 Acceptor vectors were constructed using the MSCV retroviral vector. The vectors contained in order: the left MSCV LTR (containing promoter), followed by a lox L1 sequence, followed by a hygromycin-TK fusion gene (as a selectable marker), followed by a lox L2 sequence, followed by the right MSCV LTR (see Figure 2). The retrovirus LTR was used as the promoter for the hygromycin-TK fusion gene. Similar constructs were made using other selection markers such as neomycin.

15 The L1 and L2 lox sequences of the acceptor vector had the nucleotide sequences shown below (corresponding to SEQ ID NO: 1 and SEQ ID NO: 2). L1 is the wild type LoxP1 sequence (SEQ ID NO: 1) from bacteriophage P1 (Abremski et al. (1983) *Cell* 32: 1301-1311). L2 is a mutated form of the wild type LoxP1 sequence, referred to as Lox511, having a point substitution of A for G at position 7 of the eight nucleotide spacer region (Waterhouse et al. (1993) *Nucleic Acids Res.* 21(9):2265-2266).

20

SPACER

L1 ATAACTTCGTATA ATGTATGC TATACGAAGTTAT
L2 ATAACTTCGTATA ATGTATAC TATACGAAGTTAT

25 Following construction of the acceptor vector ("L1-hygromycin-TK-L2 construct"), BOSC23 cells (ecotropic packaging cells) were transiently transfected with acceptor vector using a calcium phosphate procedure (5prime:3prime, Inc.). Viral supernatants from producers were harvested and filtered as described (Pear, *supra*, Danos, *supra*). All infections were carried out in the presence of 8 µg/ml Polybrene
30 (Sigma) Viral supernatant containing high-titer (>10⁵ pfu/ml) retroviral vector was then

used to infect host NIH 3T3 cells using the same procedures. After 48 hours in culture, the infected 3T3 cells were selected with hygromycin.

Donor vectors were constructed using pUC 19 plasmid (Yanish-Perron et al. (1985) *Gene* 33:103-119) as the backbone. The vectors contained in order: the L1 lox sequence, followed by a promoterless neomycin gene, followed by the L2 lox sequence (see Figure 2). Similar donor vectors were made using hygromycin-TK, CD24, and B-globin genes in place of the neomycin gene. Control donor vectors were constructed using a neomycin gene with the PGK (phosphoglycerol kinase) promoter, PGK-neomycin.

Various concentrations of donor vector containing neomycin gene were co-electroporated, along with an expression vector containing the Cre recombinase gene, into the infected 3T3 cells. The concentrations of donor vector ranged from 10 µg to 200 µg. After 48 hours in culture, transformed cells were selected with neomycin. Concentrations 100 µg or more of donor vector resulted in a 10-30% integration

efficiency (as measured by transfer of neomycin gene for hygromycin gene).

Different ratios of donor vector and Cre expression vector, ranging from 20:1 to 1:1 were co-electroporated into the infected 3T3 cells. All ratios resulted in the transfer of the neomycin for the hygromycin. However, a ratio of 3:1 (donor:Cre) resulted in the highest integration efficiency.

The following table provides the results of neomycin gene integration using various donor and Cre expression vectors (see Figure 3) at a concentration of 100 µg of donor vector (DNA) at a ratio of 3 parts donor vector to 1 part Cre expression vector. Experiments E#1-4 were performed as negative controls. E#5 was the positive control.

- 21 -

	Cells Electroporated	Constructs Used	# Colonies (out of 10^7 cells)
5	E#1 3T3	lox 1-PGKNeo-lox 2 & control Cre expression vector	530
	E#2 3T3	lox 1-Neo-lox 2 & control Cre expression vector	10
10	E#3 3T3	lox 1-Neo-lox 2 & Cre expression vector	2
	E#4 3T3 containing lox A-hygro-TK-lox B	lox 1-Neo-lox 2 & control Cre expression vector	21
15	E#5 same	lox 1-Neo-lox 2 & Cre expression vector	confluent ($>10^5$)

E#1 used the control donor vector (see Figure 2), lox 1-PGKNeo-lox 2 (containing the neomycin gene and a promoter) along with a control Cre expression vector (see Figure 3) (in which the sequence encoding Cre had been deleted and replaced by a gene encoding CAT). Host cells did not contain integrated acceptor vector. Therefore, E#1 demonstrated the amount of neomycin resistance conferred by random integration of the L1-PGKNeo-L2 vector capable of expressing the neomycin gene. As expected, the conferred neomycin resistance was in the range of efficiency of integration obtained by electroporation (e.g., about 0.1% efficiency).

E#2 used donor vector (promoterless) with a control Cre expression vector. Host cells did not contain integrated acceptor vector. Therefore, E#2 demonstrated the resistance conferred in the absence of acceptor vector or Cre recombinase (i.e., in the absence of efficient recombination and gene transfer). As expected, this was very low.

E#3 used donor vector (promoterless) with a functional Cre expression vector. Host cells did not contain integrated acceptor vector. Therefore, E#3 demonstrated the resistance conferred in the absence of acceptor vector (i.e., in the absence of efficient recombination and gene transfer), but in the presence of Cre recombinase. As expected, this was very low.

E#4 used donor vector (promoterless) with a control Cre expression vector (no Cre expression). Host cells contained integrated acceptor vector (L1-hygro-TK-L2). Therefore, E#4 demonstrated the gene transfer efficiency from donor vector to acceptor vector in the absence of Cre). As expected, this was very low.

E#5 used donor vector (promoterless) with a functional Cre expression vector. Host cells contained integrated acceptor vector (L1-hygro-TK-L2). Therefore, E#5 demonstrated the gene transfer efficiency from donor vector to acceptor vector in the presence of Cre. As shown in the table above, the host cells became confluent, demonstrating a greater than 1000 fold increase in gene transfer efficiency and stability.

Conclusion:

The foregoing studies and results demonstrate that the retroviral Cre/lox mediated gene transfer system of the present invention can be used for highly efficient and stable integration of transgenes into chromosomal DNA of mammalian cells.

EXAMPLE 2 - CRE/LOX MEDIATED GENE INTEGRATION USING

INVERTED LOX SEQUENCES

Donor and Acceptor Plasmids

Plasmid constructions were performed as described above in Example 1 using standard procedures. The CMV-HYTK cassette was derived from plasmid pTgCmvhytk (Immunex, Seattle, Wa). The CMV-EGFP and GFPuv cassettes were derived from plasmids pEGFP-N1 and pGFPuv (Clontech, Palo Alto, CA). The β -GFP cassette was created by replacing the CMV promoter in plasmid pEGFP-N1 by the β -globin

promoter (fragment -374 to +44 relative the cap site). Cassette HS234- β GFP was produced by linking fragment HS234 (Bouhassira et al. (1997) *Blood* 90:3332-3344) to the β -GFP cassette. Cassette HS3- β -globZ was from pCAS3 (O'Gorman et al. 1991) *Science* 251:1351-1355. A series of plasmids containing a cassette flanked by Lox sites, L1 (SEQ ID NO:1) and L2 (SEQ ID NO:2), in various orientations were created. Each plasmid was designated by the Lox site 5' of the cassette, the name of the cassette and the Lox site 3' of the cassette. Inverted L1 Lox site were designated as 1L.

The following plasmids were used in the studies described below: pL1-HYTK-L2, pL1-HYTK-1L, pL1-CMVEGFP-L2, pL1-Gfpuv-L2, pL2-GFPuv-L2, pL1-HS234 β GFP-1L, pL1-HS3 β globZ-L2EGFP (contains EGFP 3' of the L2 Lox site).

Cre Expression Plasmid (mammalian):

Plasmid pBS 185 (CMV-CRE) was obtained from Clontech (Palo Alto, CA). pSSR73 (RSV-CRE) was a gift of Dr. P. Leboulch (Harvard University, Cambridge MA), pMC-Cre was a gift of Klaus Rajewsky (Koln, Germany) (Gu et al. (1993) *Cell* 73:1155-1164).

Cre Expression Plasmid (bacterial):

A PCR generated fragment containing the coding sequence of the CRE recombinase was cloned into the NdeI/andXhoI sites of plasmid pET23c (Novagen, Madison, WI) resulting in a plasmid (pET23c-CRE) that expresses a His-tagged CRE protein in induced BL21 bacteria.

CRE Purification:

BL21 (DE3) Lys E bacteria (Novagen, Madison, WI) containing plasmid pET23c-CRE were induced with 2 mM IPTG for 2 hours, sonicated in basic buffer (50mM NaPi pH 8, 200mM NaCl) plus 0.05% Tween 20 and 1% lysozyme, centrifuged at 4000g for 10 min at 4°C, and the supernatant was loaded on a Talon metal affinity Resin column (Clontech, Palo Alto, CA). The column was washed successively with 20 ml of basic buffer containing 0, 5 and 10 mM imidazol, and the CRE protein was then

- 24 -

eluted with 5 ml of basic buffer containing 100mM imidazol. In vitro CRE recombination were performed as in ¹⁸.

Cell Culture and Electroporation:

5 Culture and DNA recombination reactions in MEL cells were performed as in (Bouhassira et al. (1997) *Blood* 90:3332-3344). AK-7 ES cells were cultured and selected on SNL feeder cells as described (Soriano P. (1997) *Development* 124:2691-2700). DNA recombination was performed by electroporating (250V, 500uF, Biorad gene pulser, Biorad, Hercules, CA) 3×10^6 cells with 200 μ g of L1GFP1L and 30 μ g of
10 MC-Cre. Selection with 3 μ M gancyclovir was applied 5 days after the transfection.

GFP Expression Studies:

5×10^5 cells were rinsed once in PBS and resuspended in Hank's solution containing 5% Fetal Calf Serum and 2 μ M propidium iodide, and analyzed on a
15 FACSCAN flow cytometer (Becton Dickinson, CA).

Site Specific DNA Integration Using Inverted Lox Sequences

The following studies were performed to determine whether the efficiency of site-specific DNA integration could be increased using two inverted lox L1 (SEQ ID
20 NO:1) sequences, in place of the lox L1 and L2 sequences used in Example 1. Thus, the L2 site was replaced with an inverted L1 site (termed 1L). As shown in Fig. 4 (panel C), intra-chromosomal recombination between two inverted Lox sites leads to inversion of the HYTK negatively selectable marker previously integrated into the genome, rather than to its excision, and therefore does not remove the sensitivity to negative selection
25 by gancyclovir. In this system, the donor cassette integrates in one orientation in half of the clones and in the reverse orientation in the other half.

A plasmid (pL1-HYTK-1L) containing the HYTK gene flanked by two inverted but identical Lox L1 (SEQ ID NO:1) sites was created, and MEL cell lines with single integrated copies of this plasmid were produced as described above. Three clones,
30 termed RL4, RL5 and RL6 were selected for further studies. Exchange reactions were performed as above using a plasmid in which inverted L1 Lox sites flank a GFP reporter

- 25 -

(pL1-HS234GFP-1L). After the transfection, 12 Gan^R clones per cell line were picked and analyzed by Southern blots. At loci RL4 and RL5, all 12 Gan^R clones tested had undergone an exchange of the HYTK cassette for the GFP cassette (Fig. 5 (panel A)). At RL6, 11 out of 12 clones tested had undergone an exchange of the HYTK cassette for the GFP cassette. As expected, about half of the clones were in reverse orientation. These results demonstrate that DNA cassette exchange with inverted lox sites can be efficiently selected using only negative selection in MEL cells.

The same experiments were then repeated in an embryonic stem cell line: Two single copy L1-HYTK-1L tagged-loci were created and tested for cassette exchange using plasmid pL1-CMVEGFP-1L. At the first locus tested, about 50% of the 60 Gan^R clones tested by Southern blot had undergone an exchange (Fig. 5 (panel B)). At the second locus the frequency of exchange was about 10%. As in MEL cells, clones in both orientations were obtained in approximately equal proportion.

15 **EXAMPLE 3 - CRE/LOX MEDIATED GENE INTEGRATION USING NO SELECTABLE MARKER**

The materials and methods used in the studies described below were as described in Examples 1 and 2 above.

Typically, site specific DNA recombination systems employ an active negative selectable marker at the integration site prior to exchange. Since, in many situations, it would be advantageous to perform recombination at loci that do not contain any active genes at all, the following system was designed which does not require the presence of an active gene before or after the recombination (e.g., cassette exchange).

When no selection is applied, about 1% of the MEL cells that survive the transfection undergo recombination. To increase this frequency, a GFP reporter gene was placed in the exchange plasmid, outside the exchange cassette, and cells that expressed the highest levels of GFP were selected for to increase the proportion of selected cells which had undergone exchange. This is because the sorted cells would be the cells that have taken up the highest amount of the exchange plasmid during the transfection (Figure 4 (panel D)).

Specifically, a plasmid containing a GFP reporter gene located outside of a LacZ exchange cassette (pL1-HS3 β globZ-L2-GFP) was created and co-transfected with a CRE expression plasmid in two lines of MEL cells (RL1 and RL3) containing pre-integrated target L1 and L2 Lox sites flanking the HYTK gene. Forty-eight hours post-transfection, the cells expressing the highest amount of GFP (0.1 percentile) were sorted individually into 96 well plates, expanded without applying any selection and tested for DNA cassette exchange. Seventy-four clones were obtained for the RL1 line and 32 for the RL3 line. The clones that had lost the HYTK gene were identified by culture in hygromycin, and the Hyg^S clones were then tested by Southern blots: At locus RL1, 4% (3/74) of the clones obtained had an exchange and 12% had lost the locus without exchange, probably because of a CRE-mediated excision. At locus RL3, 16% (5/32) of the clones had an exchange and 12% had lost the locus without exchange. This demonstrates that exchange with no expressed selectable marker in the genome can be performed at relatively high frequency by simply sorting the cells that have been transfected with a large number of exchange plasmids.

EXAMPLE 4 - SITE-SPECIFIC INTEGRATION GREATLY IMPROVES REPRODUCIBILITY OF EXPRESSION

The materials and methods used in the studies described below were as described in Examples 1 and 2 above.

To determine whether site-specific chromosomal integration leads to expression that is more reproducible and predictable than expression of the same cassette, randomly integrated. Mel cell clones with integration of the CMV-EGFP cassette at random sites were generated by transfection of plasmid pEGFP-N1, and compared with clones containing the CMV-EGFP cassette integrated by site-specific Cre/lox-mediated exchange at locus RL4.

Southern blots revealed that most of the random clones contained multiple integrated copies (Figure 5 (panel D)) and that, as expected, the clones that had undergone site-specific exchange had integration in both possible orientations. Nine random clones and nine site-specific clones with integration in one of the two possible orientations were analyzed for expression by flow cytometry (fig. 3). Mean levels of

green fluorescence for the random clones varied from 5.3 (which is similar to untransfected controls) to 357.5. The average was 94.1 ± 103.5 . Mean levels of fluorescence for the site-specific clones varied from 100.1 to 147.4. The average was 127.8 ± 11.9 . Although the average expression of the two types of clones were similar, the 10-fold difference between their standard deviations clearly demonstrates that expression levels of clones integrated at the same locus by site-specific recombination are much less variable than those of randomly integrated clones. Similar results were obtained at the 5 other targeted, site-specific loci that were tested.

The highly reproducible expression observed when site-specific clones were tested in the absence of any selective pressure at the locus suggests that controlling the integration site and therefore eliminating position-effects permits meaningful comparisons of multiple transgenes individually integrated at a particular genomic site.

Conclusion

Overall, the studies described above demonstrate that highly efficient methods of Cre/lox-mediated recombination, such as those described in the studies above, can be used to achieve site-specific recombination with no incoming selectable marker and in a defined orientation. These methods also can be used to achieve multiplex integration of transgenes (Sauer B. (1996) *Nucleic Acids Res.* **24**:4608-4613) sequentially at the same site or at multiple independent sites using multiple pairs of lox sites, either as heterospecific pairs or inverted as described herein. These methods also can be used in combination with FLP (Seibler et al. (1998) *Biochem.* **37**:6229-6234)-mediated DNA exchange systems.

Prior to the present invention, expression of experimental constructs in mammalian cells or transgenic animals was difficult to control because it is markedly influenced by position effects. This limited both the analysis of cis-DNA regulatory elements for transcription and replication and the physiological analysis of proteins expressed from transgenes. The novel compositions and methods of the present invention solve these problems by permitting the exchange of DNA (e.g., containing a negative selectable marker pre-localized on the chromosome with a transgene via a Cre-mediated double recombination between inverted or mutated, incompatible Lox sites.

Integration efficiency of close to 100% of negatively selected cells can be achieved. In addition, the present invention provides a way of achieving site-specific recombination within genomic DNA with no selection at all, except for cells that have taken up plasmid transiently.

- 5 The foregoing studies demonstrate that integration of a transgene at a given genomic site leads to reproducible expression. Therefore, the compositions (e.g., DNA vectors) and methods of the present invention can be used to develop artificial genetic loci that impart specific and reproducible regulation of transgenes in higher eukaryotes. This should facilitate the analysis of cis-regulatory DNA elements governing expression
- 10 and position effects, improve control over the physiological effects of transgenes, and accelerate the development of animal models for complex human diseases.

What is claimed is:

1. A method of achieving site-specific recombination comprising, in any order, the steps of:
 - 5 (a) providing a donor DNA comprising two inverted lox sequences;
 - (b) providing an acceptor DNA comprising the same two inverted lox sequences contained in the donor DNA; and
 - (c) contacting the donor and acceptor DNA with a recombinase.
- 10 2. The method of claim 1 further comprising the step of integrating the acceptor DNA into the genome of a cell.
3. The method of claim 1 wherein the donor DNA is present in excess of the acceptor DNA.
- 15 4. The method of claim 1 wherein the recombinase is selected from the group consisting of Cre recombinase.
5. The method of claim 1 wherein the recombinase is selected from the group consisting of Flp recombinase
- 20 6. The method of claim 4 wherein the inverted lox sequences comprise a nucleotide sequence selected from the group consisting of SEQ ID NO:1, SEQ ID NO:2 and inversions thereof.
- 25 7. The method of claim 2 wherein the acceptor DNA is integrated into the genome using a retroviral vector.
8. The method of claim 2 wherein the acceptor DNA is integrated into the genome by homologous recombination.
- 30

9. A method of achieving site-specific recombination in a selected cell comprising, in any order, the steps of:

(a) introducing into the cell an acceptor DNA which integrates into the genome of the cell, the acceptor DNA comprising two inverted lox sequences:

5 (b) introducing into the cell a donor DNA comprising the same two inverted lox sequences contained in the acceptor DNA; and

(c) introducing into the cell a recombinase which catalyzes recombination between the lox sequences.

10 10. The method of claim 9 wherein the acceptor DNA further comprises a negatively selectable marker flanked by the two inverted lox sequences.

11. The method of claim 10 further comprising the step of negatively selecting cells which have undergone site-specific recombination.

15

12. The method of claim 9 wherein the donor DNA introduced into the cell is present in excess of the acceptor DNA.

13. The method of claim 9 wherein the recombinase is introduced into the
20 cell by transfecting the cell with a gene encoding the recombinase.

14. The method of claim 9 wherein the recombinase is introduced into the cell in the form of a protein.

25 15. The method of claim 9 wherein the inverted lox sequences comprise a nucleotide sequence selected from the group consisting of SEQ ID NO:1, SEQ ID NO:2 and inversions thereof, and the recombinase is Cre.

16. The method of claim 9 wherein the donor DNA further comprises a
30 transgene encoding a therapeutic protein.

17. The method of claim 16 wherein the protein is β -globin.
18. The method of claim 9 wherein the donor DNA further comprise a positively selectable marker.
- 5 19. The method of claim 9 wherein the cell is a mammalian cell.
20. The method of claim 9 wherein the donor DNA, the acceptor DNA or both the donor DNA and the acceptor DNA are introduced into the cell by
10 microinjection into a fertilized egg.
21. The method of claim 9 wherein the donor DNA, the acceptor DNA or both the donor DNA and the acceptor DNA are introduced by microinjection into an embryonic stem cell.
- 15 22. A DNA vector comprising two inverted lox sequences.
23. The vector of claim 22 wherein the inverted lox sequences comprise a nucleotide sequence selected from the group consisting of SEQ ID NO:1, SEQ ID NO:2
20 and inversions thereof.
24. The vector of claim 22 further comprising a selectable marker.
25. The vector of claim 22 which is a retroviral vector or an adeno-associated
25 viral vector.
26. A gene transfer system comprising:
- (a) acceptor DNA which integrates into the genome of a host cell, the acceptor DNA comprising two inverted lox sequences; and
30 (b) donor DNA comprising a transgene flanked by the same two inverted lox sequences contained in the donor DNA.

27. The gene transfer system of claim 26 further comprising a recombinase which catalyzes recombination between the inverted lox sequences, or DNA encoding the recombinase.

5

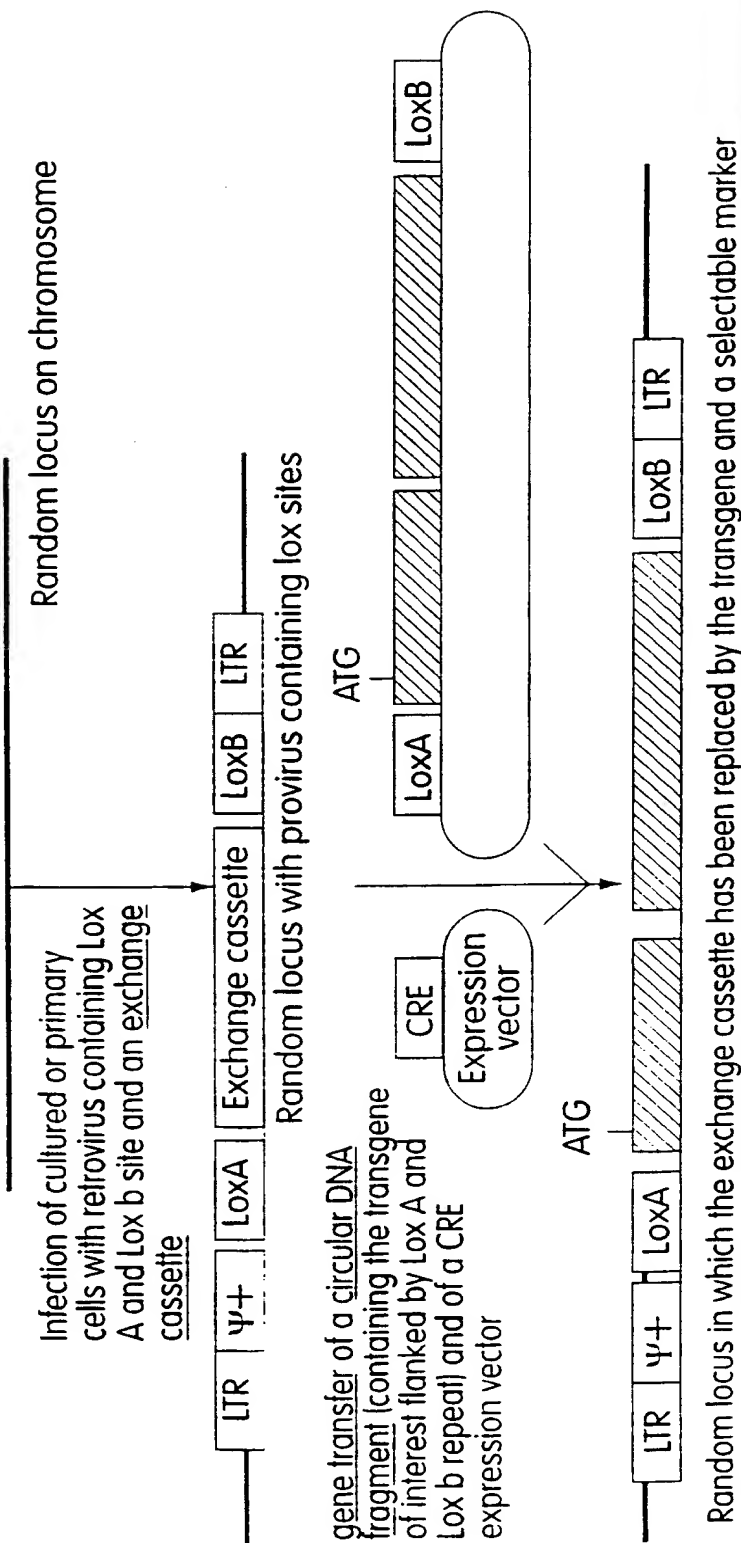
28. The gene transfer system of claim 26 wherein the acceptor DNA further comprises a negatively selectable marker.

29. The gene transfer system of claim 26 wherein the donor DNA further
10 comprises a positively selectable marker.

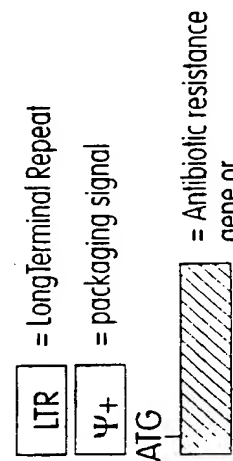
30. The gene transfer system of claim 26 wherein the inverted lox sequences comprise a nucleotide sequence selected from the group consisting of SEQ ID NO:1, SEQ ID NO:2 and inversions thereof.

15

RETROVIRUS-CRE-LOX-INTEGRATION



LEGENDS AND DEFINITIONS



= Long Terminal Repeat

= packaging signal

ATG

= Antibiotic resistance gene or gene selectable by any other means

LoxA AND LoxB = 2 mutually incompatible Lox sites

Exchange cassetts: Can be an antibiotic resistance gene or fusion gene (such as Hyrk), a gene that can be used to identified the infected cells by FACS or any other means, or an irrelevant DNA fragment

Gene transfer: can be by electroporation, lipofection, or any other means of introducing DNA into the cells

In this diagram the Cre enzyme is coligated as a plasmid. It could also be provided as a protein or be in the exchange cassette

———— = Chromosome

Circular DNA fragment: Plasmid, Cosmid or other similar vector as well as DNA fragment circularized by enzymatic reactions

Fig. 1

2/10

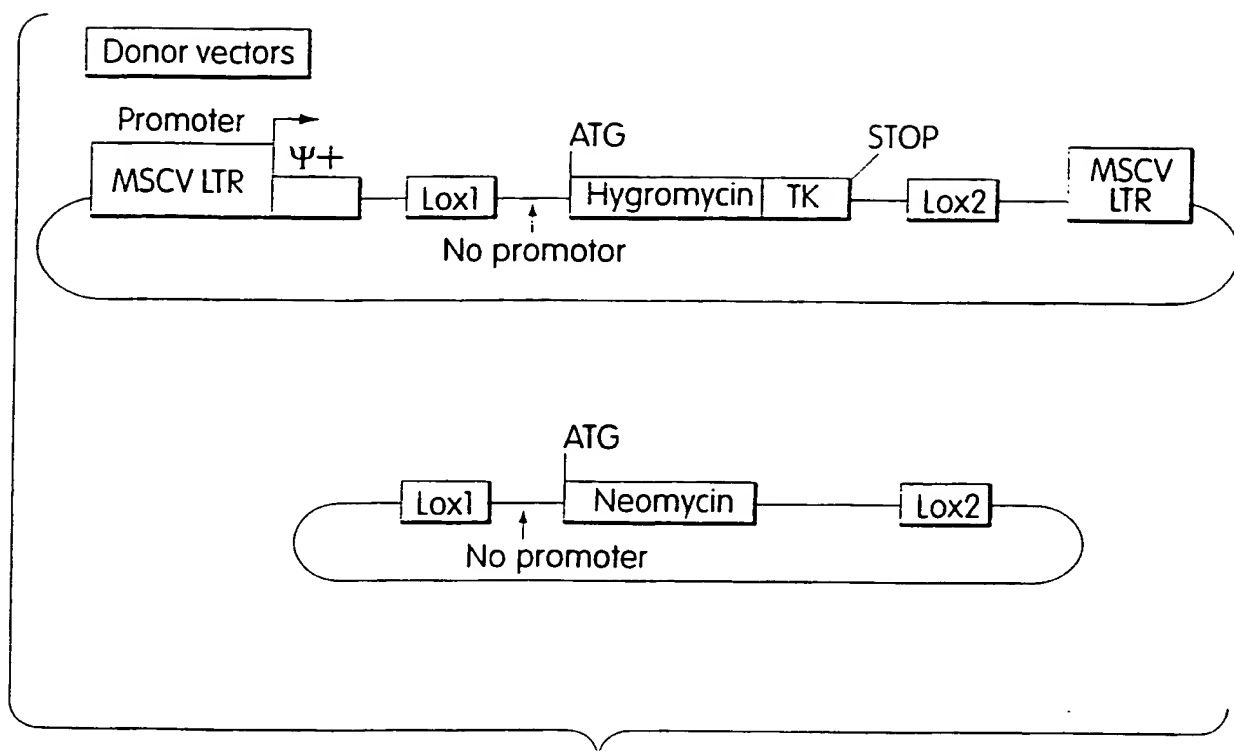


Fig. 2

3/10

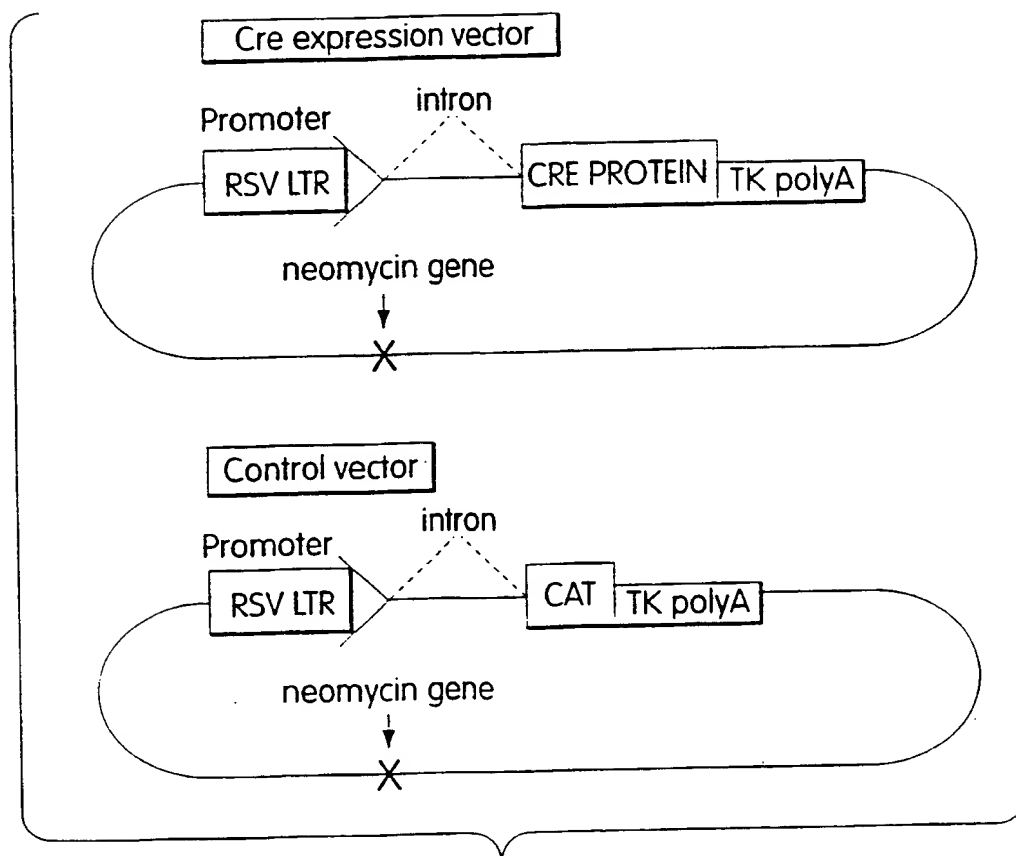
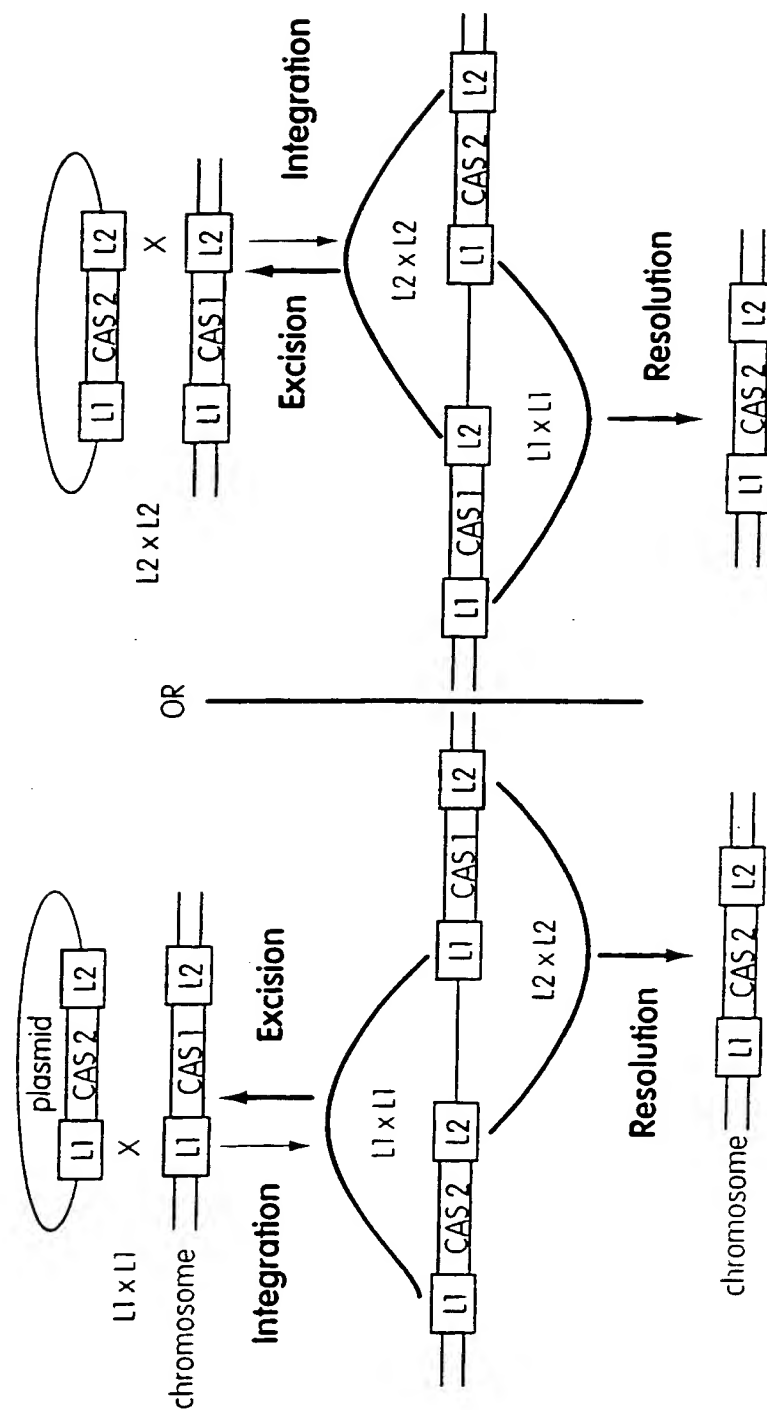


Fig. 3

4/10



PRINCIPLE OF RMCE

Fig. 4A

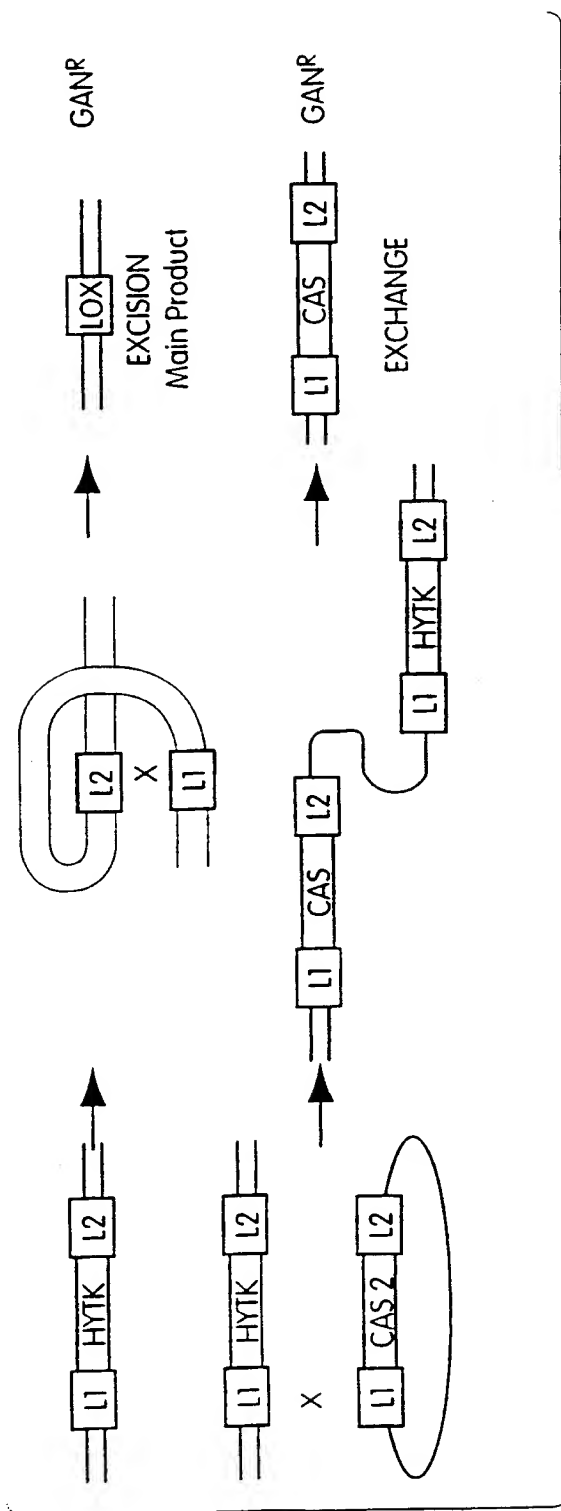


Fig. 4B

6/10

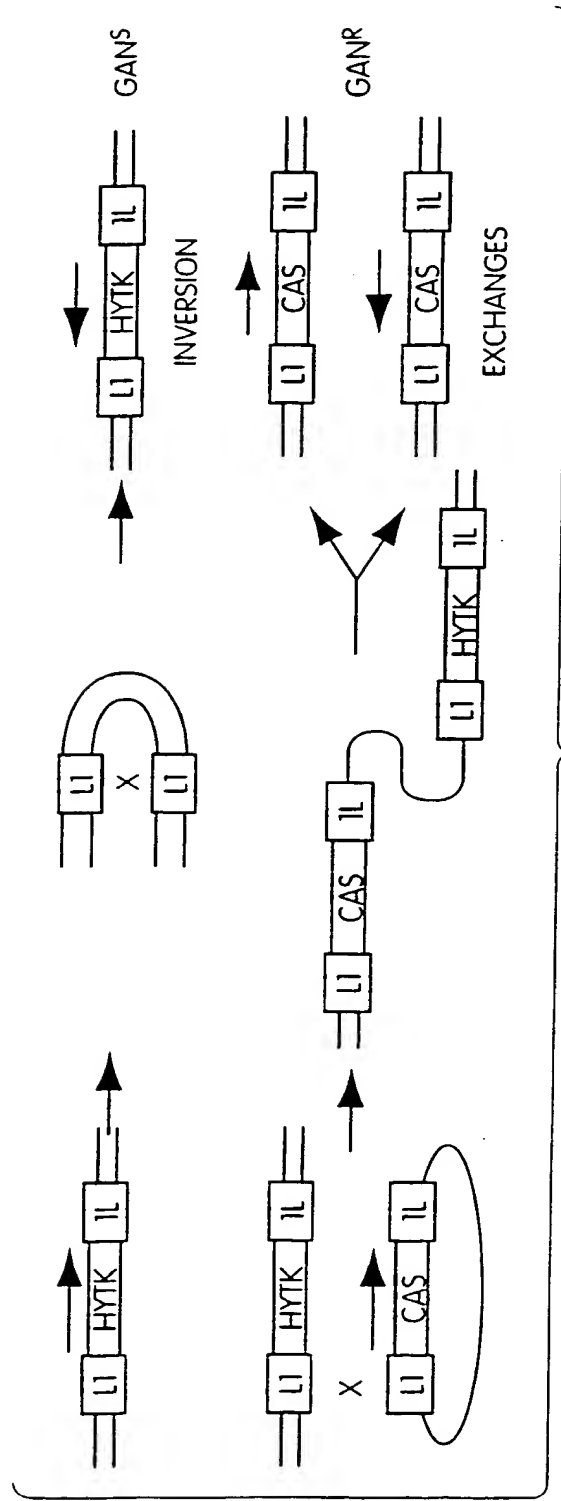
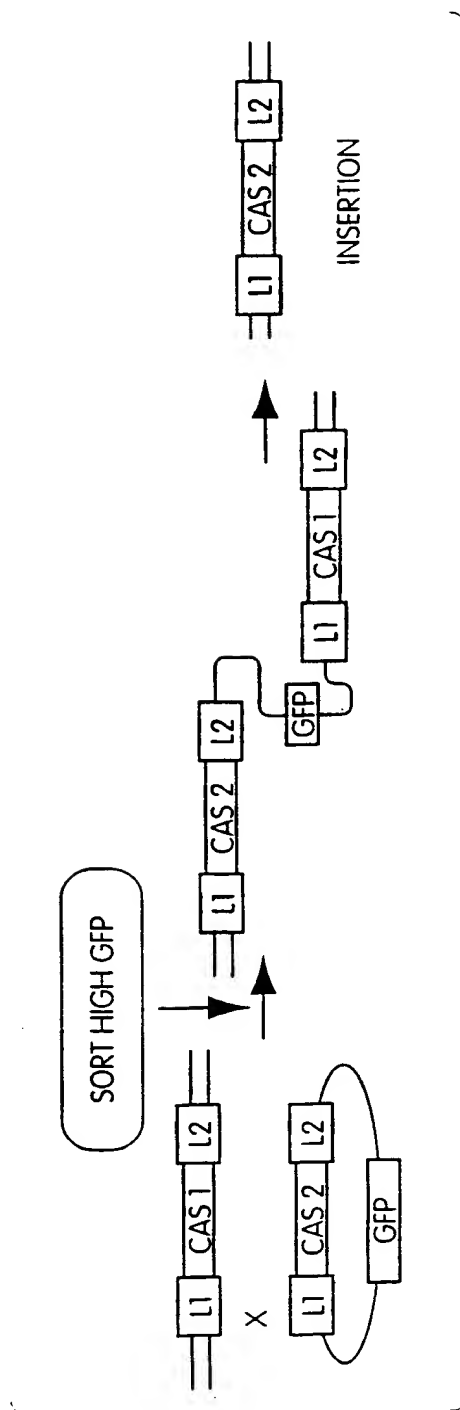


Fig. 4C

7/10



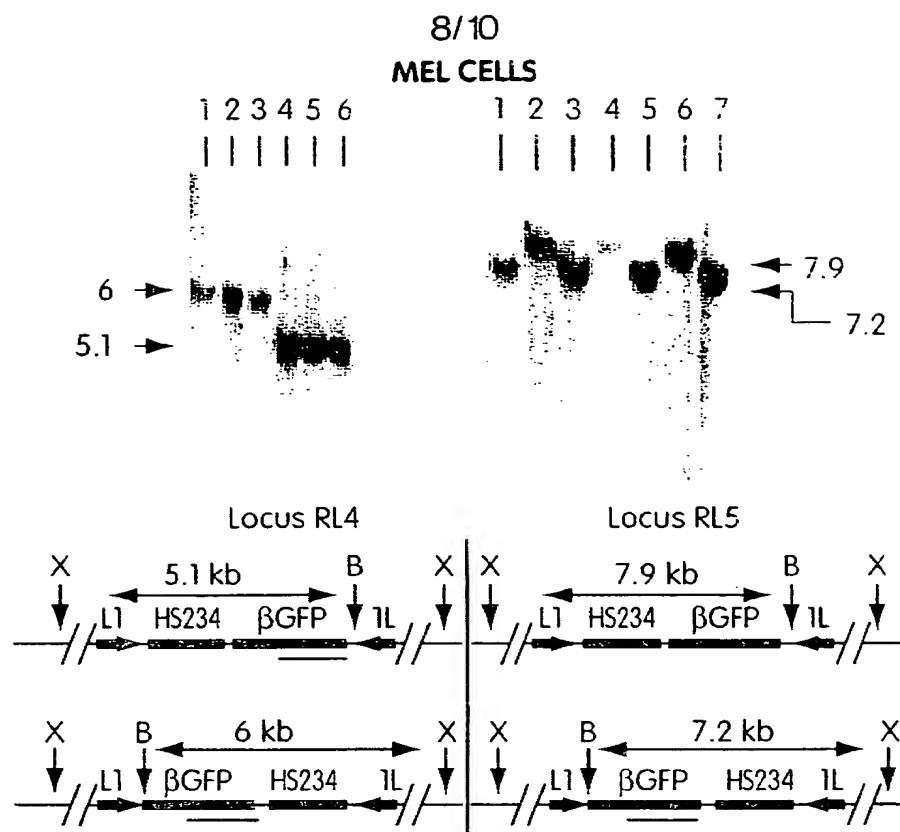


Fig. 5A

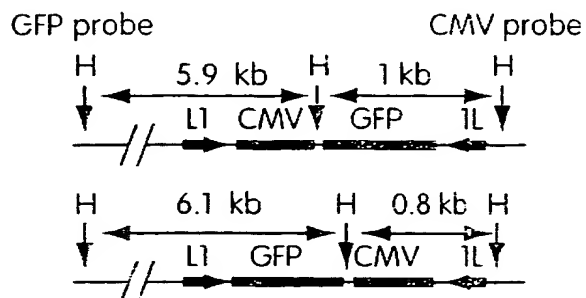
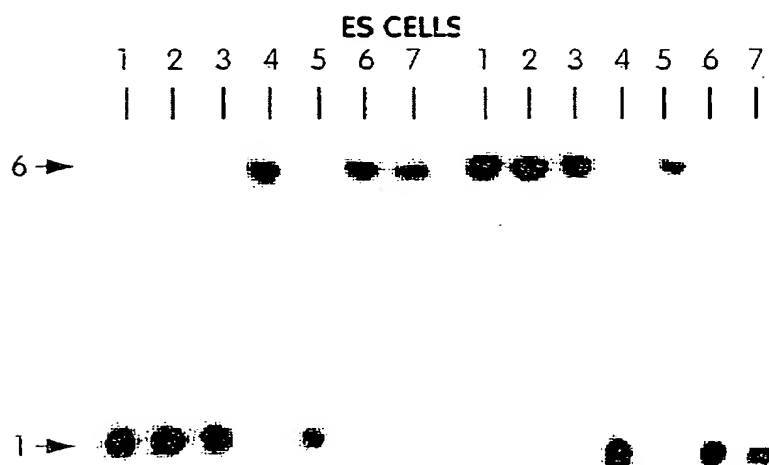


Fig. 5B

SUBSTITUTE SHEET (RULE 26)

9/10

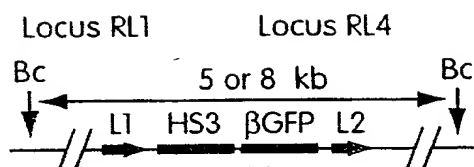


Fig. 5C

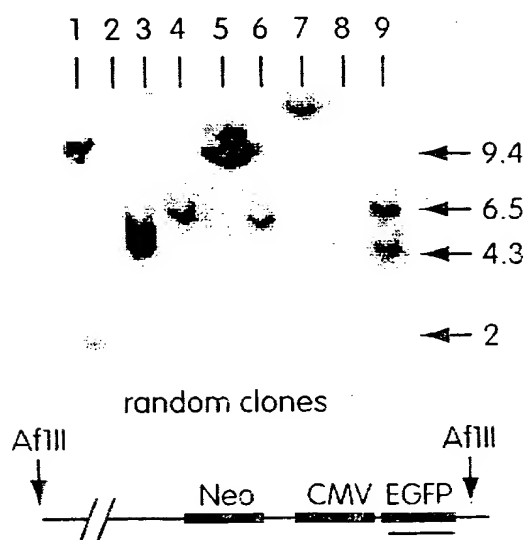


Fig. 5D

10/10

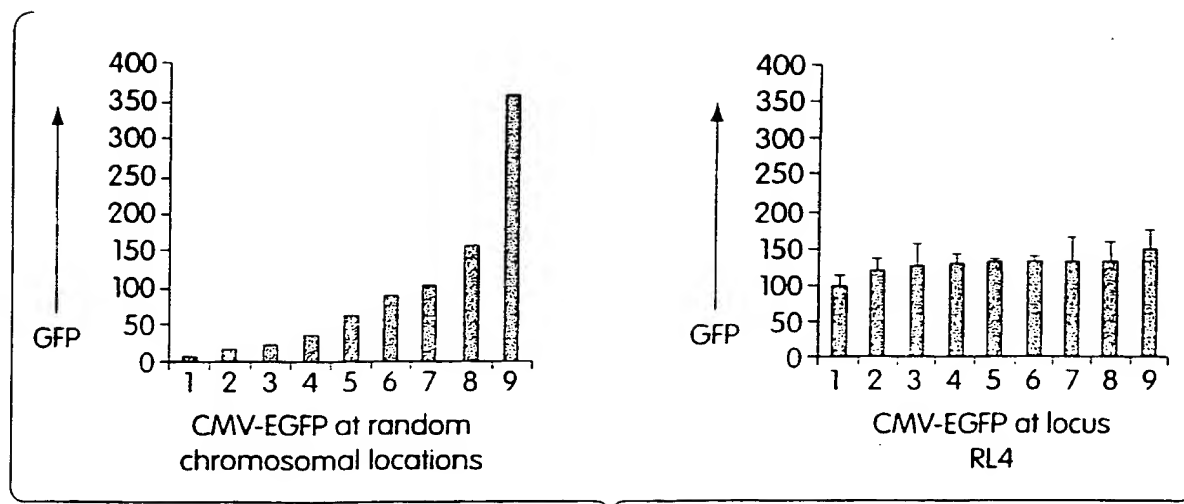


Fig. 6

- 1 -

SEQUENCE LISTING

<110> MASSACHUSETTS INSTITUTE OF TECHNOLOGY and
ALBERT EINSTEIN COLLEGE OF MEDICINE OF YESHIVA UNIVERSITY

<120> METHODS AND COMPOSITIONS FOR TRANSFORMING CELLS

<130> MTZ-198CPPC

<140> 09/293,303

<141> 1999-04-16

<160> 2

<170> PatentIn Ver. 2.0

<210> 1

<211> 34

<212> DNA

<213> Artificial Sequence

<220>

<223> Description of Artificial Sequence:integration site

<400> 1

ataacttcgt ataatgtatg ctatacgaag ttat

34

<210> 2

<211> 34

<212> DNA

<213> Artificial Sequence

<220>

<223> Description of Artificial Sequence:integration site

<400> 2

ataacttcgt ataatgtata ctatacgaag ttat

34

INTERNATIONAL SEARCH REPORT

International Application No

PCT/US 00/09782

A. CLASSIFICATION OF SUBJECT MATTER

IPC 7 C12N15/90

According to International Patent Classification (IPC) or to both national classification and IPC

B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)

IPC 7 C12N

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched

Electronic data base consulted during the international search (name of data base and, where practical, search terms used)

BIOSIS

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category *	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
Y	WO 97 47758 A (EINSTEIN COLL MED ; MASSACHUSETTS INST TECHNOLOGY (US)) 18 December 1997 (1997-12-18) the whole document	1-30
Y	BOUHASSIRA E. E. ET AL.: "Transcriptional behavior of LCR enhancer elements integrated at the same chromosomal locus by recombinase-mediated cassette exchange." BLOOD, vol. 90, no. 9, 1997, pages 3332-3344, XP002143981 ISSN: 0006-4971 figure 1	1-30



Further documents are listed in the continuation of box C.



Patent family members are listed in annex.

* Special categories of cited documents :

- "A" document defining the general state of the art which is not considered to be of particular relevance
- "E" earlier document but published on or after the international filing date
- "L" document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)
- "O" document referring to an oral disclosure, use, exhibition or other means
- "P" document published prior to the international filing date but later than the priority date claimed

"T" later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention

"X" document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone

"Y" document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art.

"&" document member of the same patent family

Date of the actual completion of the international search

1 August 2000

Date of mailing of the international search report

21/08/2000

Name and mailing address of the ISA

European Patent Office, P.B. 5818 Patentlaan 2
NL - 2280 HV Rijswijk
Tel. (+31-70) 340-2040, Tx. 31 651 epo nl.
Fax: (+31-70) 340-3016

Authorized officer

Mandl, B

INTERNATIONAL SEARCH REPORT

Inte. onal Application No

PCT/US 00/09782

C.(Continuation) DOCUMENTS CONSIDERED TO BE RELEVANT

Category *	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
Y	FR 2 745 008 A (ASS POUR LE DEV DE LA RECH EN) 22 August 1997 (1997-08-22) figure 1 -----	1-30
Y	WO 97 09439 A (GENVEC INC ;KOVESDI IMRE (US); MCVEY DUNCAN L (US)) 13 March 1997 (1997-03-13) page 7, line 19 -page 11, line 8 -----	1-30
P,X	FENG Y.-Q. ET AL.: "Site-specific chromosomal integration in mammalian cells: Highly efficient CRE recombinase-mediated cassette exchange." JOURNAL OF MOLECULAR BIOLOGY, vol. 292, no. 4, 1 October 1999 (1999-10-01), pages 779-785, XP000929622 ISSN: 0022-2836 the whole document -----	1-30

INTERNATIONAL SEARCH REPORT

Information on patent family members

International Application No

PCT/US 00/09782

Patent document cited in search report	Publication date	Patent family member(s)	Publication date
WO 9747758 A	18-12-1997	US 5928914 A	27-07-1999
		AU 3306297 A	07-01-1998
		CA 2258007 A	18-12-1997
		EP 0914457 A	12-05-1999
FR 2745008 A	22-08-1997	AU 707684 B	15-07-1999
		AU 2098997 A	10-09-1997
		CA 2247517 A	28-08-1997
		EP 0896620 A	17-02-1999
		WO 9731108 A	28-08-1997
		JP 2000505298 T	09-05-2000
WO 9709439 A	13-03-1997	US 5801030 A	01-09-1998
		AU 717597 B	30-03-2000
		AU 6912296 A	27-03-1997
		BG 102344 A	30-04-1999
		CN 1200149 A	25-11-1998
		CZ 9800612 A	15-07-1998
		EP 0850312 A	01-07-1998
		HU 9802538 A	01-02-1999
		NO 980838 A	21-04-1998
		NZ 316806 A	28-10-1999
		PL 325372 A	20-07-1998
		SK 27798 A	09-09-1998
		US 6063627 A	16-05-2000

THIS PAGE BLANK (USPTO)